Part 1 - Session Papers for the EPA 22nd Annual National Conference on Managing Environmental Quality Systems

April 14-17, 2003 New Orleans, Louisiana

OPENING PLENARY

Evaluation of Effectiveness of Bioassays and Remediation Strategy for Agent Orange Contaminated Soil and Sediments at the Former Danang Air Base in Vietnam * - Vance S. Fong, P.E., U.S. EPA

IMPROVING STATISTICAL METHODOLOGIES

A Statistical Perspective on Recent Developments in Chemical Analytical Detection and Measurement * - H. Kahn, U.S. EPA

What is your IQ? * - Cynthia Curtis, Kevin Bolger, U.S. EPA

AMBIENT AIR

Ozone Data Quality Objective Development - Basil Coutant, Battelle Memorial Institute

PM Coarse Data Quality Objective Tool Development - Basil Coutant, Battelle Memorial Institute

The QA Strategy Workgroup's Validation Templates * - Tom Parsons, South Coast Air Quality Management District

QUALITY SYSTEMS CONCEPTS, MODELS, AND METHODOLOGIES

Driving Excellence - Quality Methods at the National Vehicle and Fuels Emissions Laboratory - James Carpenter, U.S. EPA, Thomas Schrodt, U.S. EPA

Quality Assurance or Quality Assistance * - Lou Blume, U.S. EPA

An Intra-active Web Based Research Plan/Quality Assurance Plan – Allan R. Batterman, U.S. EPA

PLANNING AND ASSESSING QUALITY

Planning for Quality in EPA's Next Generation Mobile Source Emission Model - Chad Bailey, U.S. EPA

Systematic Planning for the Residual Risk Program: Adapting the Quality Planning Process to the Needs of the Participants * - Malcolm Bertoni, Research Triangle Institute

Practical Estimators of PM2.5 Data Quality Objectives Parameters - Basil Coutant, Battelle Memorial Institute

Evaluation and Use of Secondary Data - James T. Markwiese, Neptune and Company

An Information Quality Life Cycle Model: Extending the Reach of Quality Management to Secondary Users of Data * - Malcolm Bertoni, Research Triangle Institute

WIN-WIN-WIN PARTNERSHIP FOR TRAINING ENVIRONMENTAL STATISTICIANS: A PANEL DISCUSSION *

* **Abstract Only**

Evaluation of Effectiveness of Bioassays and Remediation Strategy for Agent Orange Contaminated Soil and Sediments at the Former Danang Air Base in Vietnam U.S.-Vietnamese Cooperative Research

Vance S. Fong, P.E., U.S. EPA, Quality Assurance Manager

Introduction

One of the major areas of cooperation signified in the July 2001 and the March 2002 U.S. – Vietnam Dioxin Research Agreements is to develop capacity to identify areas with lingering high levels of dioxin in Vietnam. "Hot spots" containing high levels of dioxin in soil have been identified and others are presumed to exist but have yet to be located. These “hot spots” are potentially associated with former bases occupied by U.S. forces and with areas defoliated during the Vietnam War.

Both countries expressed support for an environmental survey to identify potential populations for health studies and “hot spots” for possible remediation. The overall goal of this collaboration is to advance science for the purpose of reducing human exposure and improving public health. Research is needed to develop approaches for more rapid and less expensive screening of dioxin residue levels in soil, sediments, and biological samples which can be applied in Vietnam. These approaches can then be used to more readily locate highly contaminated areas, to monitor effectiveness of remediation, and to understand migration of dioxin in the natural environment. Monitoring efforts need to be linked to modeling efforts to understand fate and transport of dioxin in the environment. The overall strategy also includes evaluation of innovative and cost-effective approaches to environmental remediation for application in Vietnam.

The environmental agencies of both countries strongly support the need for high quality research and capacity building as a means to identify, characterize and mitigate dioxin health, and ecological impacts, and bridge knowledge gaps regarding human exposure both the past and into the future. Coordination between health and environmental efforts will be necessary to achieve success. This has been a research collaboration among the Vietnam Ministry of Science, Technology and the Environment, Vietnam National Center for Natural Sciences and Technologies, U.S. National Institute of Health, U.S. EPA, U.S. Center Disease Control, and the State Department.

Progress to Date: Technology Transfer, Research, and Data Development

This is an on-going effort and is too soon to predict what the eventual results and benefits will be. Preliminary accomplishments include:

I. Trained Vietnamese scientists in sample collection and bioassay analytical techniques.
II. Planned out pilot "hot spot" assessment using effective site characterization tools.
III. Built laboratory capacity/expertise in analysis and remediation.
IV. Collected and analyzed the first set of soil samples from a "hot spots" at the former Danang Air Base.
Developed a matrix of potentially applicable remediation technologies for further evaluation.

A Statistical Perspective on Recent Developments in Chemical Analytical Detection and Measurement

Henry D. Kahn
Statistics and Analytical Support Branch
Engineering and Analysis Division
Office of Science and Technology
Office of Water
US Environmental Protection Agency

This presentation will review recent developments in chemical analytical detection and measurement related to ongoing work conducted by EPA's Office of Water. The Office is currently working to fulfill the requirements of a Court ordered settlement agreement to perform an assessment of methodology for determining levels for detection and measurement, or quantification, of chemical analytes These levels define important characteristics of techniques used to measure chemical analytes and are often controversial. The basic elements of detection and measurement will be reviewed and issues addressed in the settlement agreement will be discussed. The presentation will emphasize statistical concepts and perspective on the development and application of methodology for determining detection and measurement levels.

What is your IQ?

Cynthia Curtis, U.S. EPA and Kevin Bolger, U.S. EPA

Information quality (IQ) is an essential building block for making sound environmental decisions. Unfortunately, quality assurance information associated with data is often overlooked or simply not included with data sets. What if it was simple to retrieve pertinent QA information and its associated data? What if you only had to enter the information into a database once? What if it could be done using an on-line system that leads you through a series of simple questions? During our presentation, we will tour a pilot of an on-line system that offers these features and more. The Region 5 State of Environment (R5 SOE) QA system was developed by U.S. EPA in response to needs by regions to develop a consistent platform on which data and supporting information for multi-media analysis can be assessed, documented, and qualified. We will further describe how R5 SOE QA works and which areas we expect to enhance in the future.

Ozone Data Quality Objective Development

Basil Coutant, Battelle

Development of the Data Quality Objectives (DQOs) for the 8-hour ozone standard is being completed. The process currently underway started with an assessment of the DQO process initiated in 1997 by EPA's Office of Air Quality Planning and Standards. That assessment identified several issues with the initial DQO development that are being addressed in the current effort.

The current DQO development uses the hourly ozone concentrations reported in AIRS for the three-year period of 1999-2001. These data are being analyzed to provide a distribution of parameter estimates needed for simulation models and to validate those models. The proposed simulation models are similar to the ones used for the $PM_{2.5}$ DQO development and incorporate seasonality, diurnal variability, data incompleteness, natural variability, autocorrelation, and measurement error.

The primary goal of this work is the validation of the simulation models and establishment of the national distribution of the parameters to be used in the models. Various scenarios are presented so that EPA and decision makers can make the final choices with respect to the level of uncertainty that is acceptable. Those choices will complete the DQO development process.

1.0 INTRODUCTION

The primary National Ambient Air Quality Standard (NAAQS) for ozone is based on an average from three consecutive years of the fourth highest daily value of rolling 8-hour average hourly concentrations (see 40CFR50). The Data Quality Objective (DQO) process for the attainment/non-attainment decision of the ozone NAAQS was initiated in the mid-1990s but was not completed. This report discusses the recent progress made and the steps that have been taken to complete the DQO process.

The current DQO development is based on a pair of simulation models, one of which is similar to the one used for $PM_{2.5}$ (see U.S. EPA's DQO *Companion*, Version 1.0 User's Guide). Namely, they assume that the maximum 8-hour averages follow a long-term sinusoidal pattern with random day-to-day deviations from that pattern. The daily values are also assumed to have a consistent multiplicative bias and random measurement imprecision. The sinusoidal pattern and degree (and nature) of random day-to-day scatter about the sinusoidal pattern have been estimated from ozone data obtained from AQS. The second simulation model mimics the hourly measurements taken within a day. Again, a sinusoidal pattern is assumed to mimic the diurnal patterns observed in the data. Together, these are used to study the effect of measurement inaccuracy and data incompleteness on the estimate of the 3-year mean fourth highest value and the subsequent decision errors.

2.0 DISCUSSION

This section discusses the main issues that are being incorporated into the current framework that differ from the previous ozone DQO effort or the $PM_{2.5}$ DQO development. The main issues with the development of the ozone DQOs are:

1. The database used to establish the DQO parameters.
2. The distribution used in the simulations.
3. The inclusion of autocorrelation in the simulations.
4. The establishment of a relationship between the precision and bias of the daily aggregate to the actual measurement precision and bias.
5. The use of data incompleteness in the DQO simulations.
6. The use of year-to-year variation in the simulations.

Each of these issues is discussed in the following sections.

2.1 The Data

The database for this effort is an archived collection of hourly ozone concentrations reported in AQS for the years 1999, 2000, and 2001 extracted on September 23, 2002, and available at http://www.epa.gov/ttn/airs/airsaqs/archived%20data/archivedaqsdata.htm. This database allows the parameters used in developing the DQOs to be estimated from a nationally representative set of data. "Representative" in this case does not just mean representative of the most typical data, but additionally the range of conditions that could reasonably be expected from the program. Obviously, this would have been difficult to obtain prior to the implementation of a national network. However, even if the DQOs had been completed earlier, it would be appropriate to review the assumptions made and the parameter estimates with the current 3-year database.

2.2 Choice of Distribution

For the $PM_{2.5}$ DQO development, lognormal deviations from the sinusoidal seasonal mean were used. However, the Weibull distribution has been described as better than the lognormal for estimating the extremes in the ozone data (Curran and Frank, 1975). This claim was investigated for sites with the greatest day-to-day variability, where the differences between the distributions would make the greatest difference.

Figure 1 shows three non-negative distributions each with the same mean and variance and, hence, the same coefficient of variation (CV). The distributions are very similar and have all been used to describe pollution concentrations. For the purpose of describing typical or mean behavior, it often makes little difference which is used. However, for describing extreme behavior as required for the ozone DQO, it does make a difference. Notice that the 98.9th percentiles of the three distributions differ by more than 10 percent. Hence, if we were to simulate the data using the mean and CV in the example, the fourth highest values (out of 365) would differ by more than 10 percent among these three distributions. Hence, the simulations could introduce a bias.

Figure 2 shows quantile-quantile (Q-Q) plots for the lognormal and Wiebull distributions in Figure 1. The lognormal distribution is used as the reference distribution on the horizontal axis in Figure 2 since the lognormal distribution was used in the simulations for the previous ozone DQO development and in the $PM_{2.5}$ DQO development. The plots show a small range of CV values, namely 53-60 percent. The vertical range on the right-hand-side of the plots shows the range of the 98.9th percentile for three distributions. The points indicated by x's on the far right show an adjusted mean 4th highest concentration for the sites in the database that contained at least 360 values in each of 1999, 2000, and 2001, and had root-mean-square CV's between 53 percent and 60 percent. The values for the data are all adjusted (by multiplication) so that the adjusted data for each site would have a mean of 0.026 ppm, just as the reference distribution has. Notice that the range of 98.9th percentile values for the lognormal and Wiebull distributions are disjoint and the data values are either in the range for the Wiebull distribution or below that range. This does not necessarily indicate that the Wiebull distribution is the more appropriate distribution, since the data (more specifically the CV's) have not been adjusted for seasonality. However, it does show that there is a significant difference between the distributions in a range of interest.

Given the magnitude of this potential bias, the current DQO development is using the Wiebull distribution to simulate the daily maximum concentrations. This choice should regularly be reviewed.

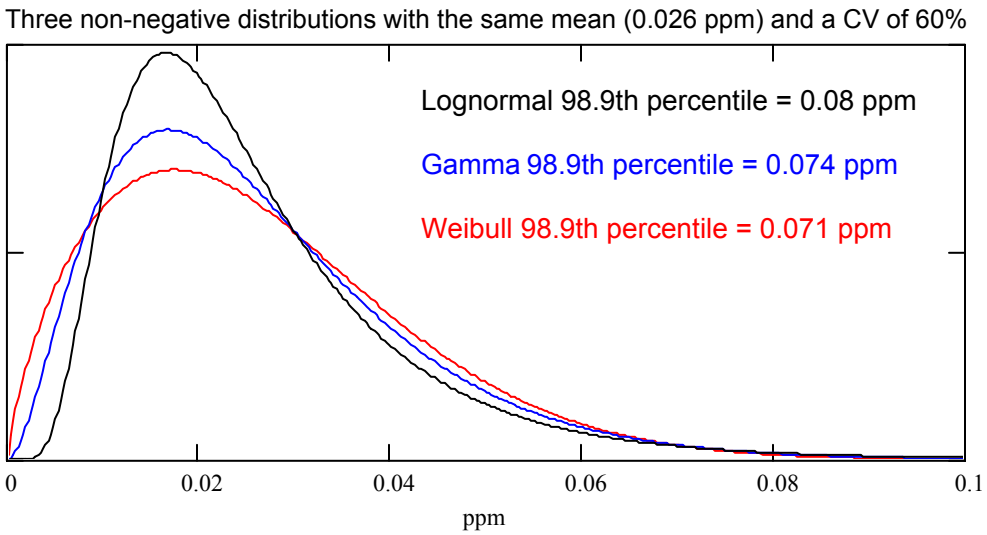

Figure 1. Three non-negative distributions with the same mean (0.026 ppm) and Coefficient of Variation (60 percent).

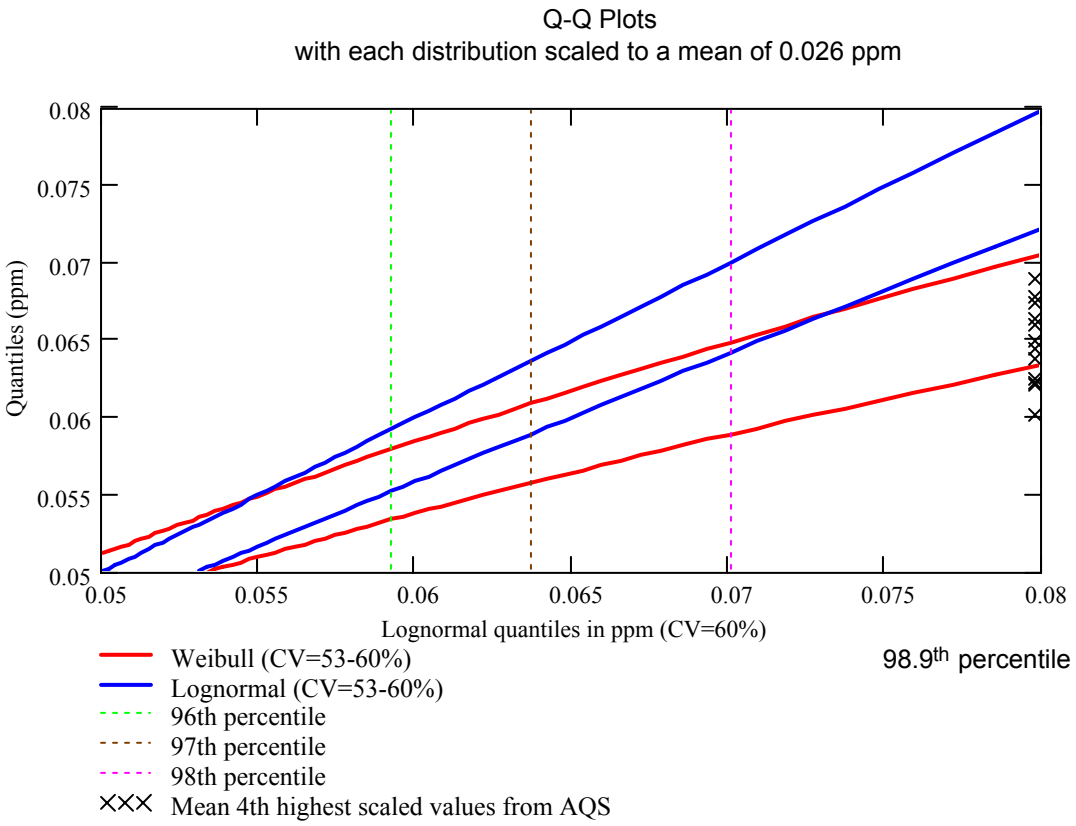

Figure 2. Q-Q plot of the example distributions and the 98.9th percentile, 3-year mean fourth highest values from 12 sites with CV's between 53 and 60 percent adjusted to have the same mean as the reference distribution.

2.3 Autocorrelation

Autocorrelation is a tendency of successive measurements to be similar. This property of the data is an assumption often used in predicting high ozone days. Generally, autocorrelation can occur for a wide variety of physical properties of a measurement process. For ozone, the primary cause would most likely be that the ozone measured at any given time has been produced over a much longer time interval than the sampling intervals. As a result, the value from one day does have some predictive value for estimating the following day's maximum 8-hour average concentration. Of course, predictive models should include other factors, such as meteorological factors and/or day of the week effects. It was felt essential to include this property as an option in the simulations. However, it may turn out that the most conservative estimate of the autocorrelation after correcting for seasonality is 0. Also, since the correlation may be a function of concentration level, it may be appropriate to estimate the autocorrelation from a restricted data set.

2.4 Translation of 8-hour Daily Maximum Precision and Bias into Measurement Precision and Bias

The precision and bias of the maximum 8-hour average is not the same as the precision and bias of the individual measurements. Assuming independence of the measurement errors and no diurnal pattern, an 8-hour average should be more precise than the individual hourly measurements by a factor of $\sqrt{8}$ or 2.83. However, the maximum 8-hour average is based on the maximum of all the 8-hour averages and virtually all areas have a diurnal pattern, so this is an over simplification. Hence, a secondary simulation process is being used to translate the desired levels of precision and bias for the daily maximums to the precision and bias requirements of the hourly measurements.

2.5 Data Completeness

Data incompleteness is incorporated in the model because data completeness has a strong effect on the ability to accurately estimate the upper percentiles of a distribution. (This can be observed with the $PM_{2.5}$ DQO tool.) While some data incompleteness should be included in the simulations used to develop the DQOs, the current data show that most sites are much more complete than the 75 percent required by CFR. Hence, it may be appropriate to use a data completeness that is representative of what is currently being achieved. Otherwise, the resulting DQOs may be unnecessarily restrictive.

Data incompleteness is also being incorporated into the within-day simulations. A small amount of missing data at the hourly level is fairly common. The simulations have shown that incompleteness on the hourly level can introduce a small positive bias in the maximum 8-hour mean for a day.

2.6 Year-to-year variation

Ozone concentrations are known to be affected by meteorological conditions. Hence, a small amount of year-to-year variation has been incorporated into the model. Figure 3 shows the distribution of the annual fourth highest values for the years 1999, 2000, and 2001 from sites with data in each of the three years. Note that the median for 1999 is approximately the 75th percentile for the other two years. The amount of variability used in the simulations is based on estimating the year-to-year effect on a site level and then using approximately the 90th percentile of those values, which is about a 6 percent coefficient of variation in the year-to-year means.

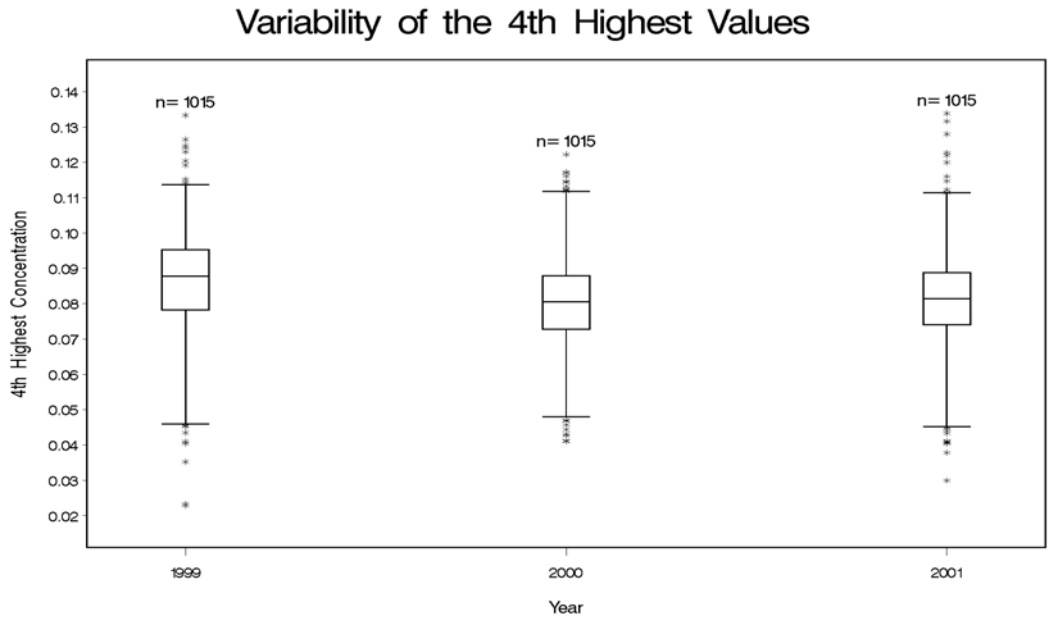

Figure 3. Box plots of the annual fourth highest values from sites with data in each of 1999, 2000, and 2001.

3.0 NEXT STEPS

The simulation model has been developed and the parameters estimated for sites with data from the 1999 though 2001 in AQS. From the site estimates, a range of values will be chosen to reflect the more extreme conditions across the nation. Finally, specific scenarios will be chosen to illustrate how different levels of precision and bias affect the attainment/non-attainment decision. Decision makers can then use these to set the precision and bias standards for ozone and complete the DQO process.

4.0 REFERENCES

Curran, T., and Frank, N. (1975). "Assessing the Validity of the Lognormal Model When Predicting Maximum Air Pollution Concentrations," *Proceedings of the 68th Annual Meeting of the Air Pollution Control Association*, June 15-20.

Code of Federal Regulations (July 1997), 40CFR50, downloaded from http://www.access.gpo.gov/nara/cfr/index.html accessed November 6, 2002.

U.S. EPA (2000). "Guidance for the Data Quality Objectives Process (EPA QA/G-4)." Report No. EPA/600/R-96/055, August.

U.S. EPA (2002). "DQO *Companion*, Version 1.0 User's Guide," written by Battelle for U.S. EPA under Contract No. 68-D-98-030, Work Assignment 5-07.

PM_{Coarse} Data Quality Objective Tool Development

Basil Coutant and Chris Holloman, Battelle

PM_{coarse} is usually measured by taking the difference between a PM_{10} measurement and a co-located $PM_{2.5}$ measurement. Hence, there are two sources of measurement error. A software tool is being developed to guide the Data Quality Objective (DQO) process for the forthcoming PM_{coarse} National Ambient Air Quality Standard (NAAQS). The tool may also be used in the decision process for establishing the percentile used for the daily NAAQS if potential data quality issues are considered.

Like the software developed for $PM_{2.5}$, this software will establish decision performance curves from a simulation model. In fact, since the measurements rely on a $PM_{2.5}$ measurement, the $PM_{2.5}$ simulation model will be directly incorporated in the PM_{coarse} simulation model. The added complexity of the measurement process is reflected in the new model. For example, two sets of ambient characteristic need to be entered; one for the PM_{10} ambient characteristics and one for the $PM_{2.5}$ ambient characteristics. Likewise, measurement characteristics of both cut-points are needed as inputs to the software.

Concurrently, with the software development, there is an assessment of PM_{coarse} measurements to estimate the likely ranges of the input parameters. These will be presented along with a demonstration of a beta version of the software.

1.0 INTRODUCTION

The National Ambient Air Quality Standards (NAAQS) for PM_{coarse} are expected to be similar to the standard for $PM_{2.5}$ with one standard based on the mean of three consecutive annual means and one based on the mean of three consecutive annual observed percentiles. However, the specific percentile has not been determined yet. To aid in the development of the standards and the Data Quality Objectives (DQOs), a simulation tool is being developed (see U.S. EPA's DQO *Companion*, Version 1.0 User's Guide). The tool simulates both the true $PM_{2.5}$ and PM_{10} concentrations and the corresponding measured values. Input parameters to the tool control aspects of the simulated ambient conditions such as the degree of seasonality for the two fractions, aspects of simulated measurement quality such as the sampling completeness, and aspects of the simulated decision process such as which percentile to use as the daily standard.

Concurrently, with the tool development, a nationwide evaluation of PM_{coarse} data has been conducted. AQS data were used to find site-specific estimations of the parameters used in the simulation models. The purpose is to understand the range of conditions that must be considered in the national level DQO development, not to develop site-specific DQOs.

2.0 SIMULATION PARAMETERS

This section describes each of the input parameters. If the input parameter describes a characteristic of the ambient conditions, then the subsection also includes a method for estimating the parameter from local measurements as well as selected quantiles from using the described method to estimate the parameter at 502 sites with co-located PM_{10} and $PM_{2.5}$ in the AQS database for the years 1999 through 2001.

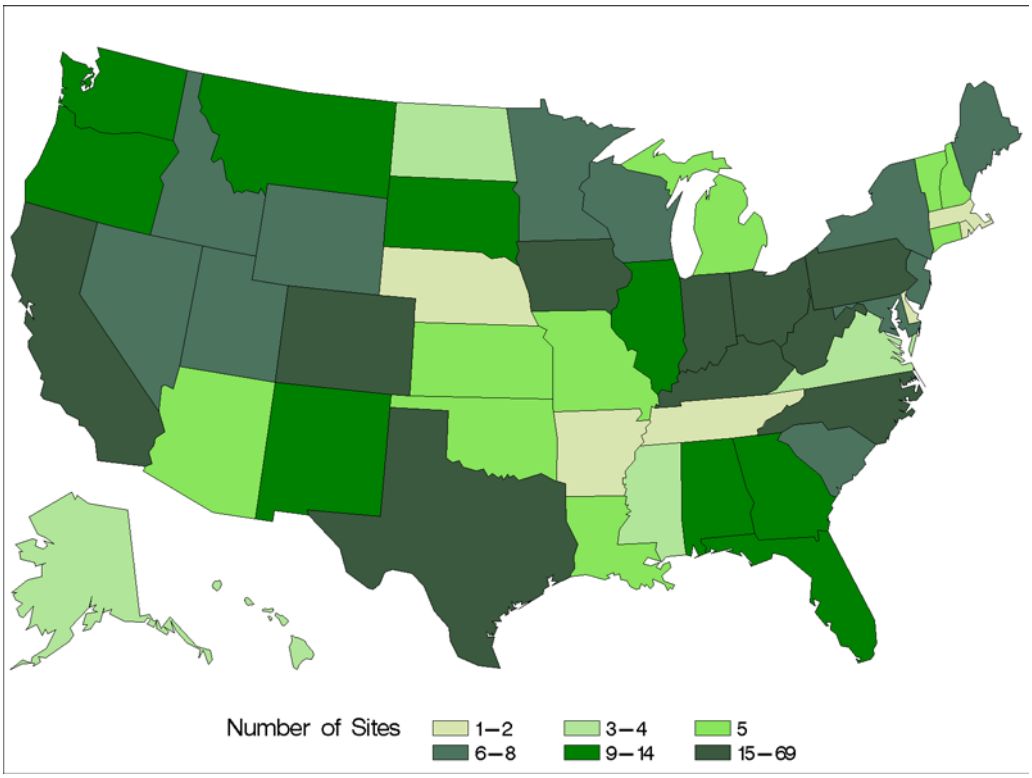

Figure 1. Distribution of the Sites Used to Estimate the Simulation Parameters

2.1 Ratio

The ratio parameter is a measure of the degree of seasonality in the data. It is the ratio of the high point to the low point on the sine curve that describes the long-term average behavior of PM. This ratio must be estimated separately for the $PM_{2.5}$ and PM_{coarse} series. With at least a year of data, the ratio is estimated by calculating the means for each month and dividing the highest monthly mean by the lowest monthly mean. For sites with more than 1 year of data, the monthly averages are based on all of the data for the given month even though they may come from different years. Table 1 shows quantiles of the estimated ratio parameter for $PM_{2.5}$ and PM_{coarse} across several sites in the United States.

Table 1. Seasonality Ratio Quantiles

Quantile	2.5	10	20	30	40	50	60	70	80	90	97.5
$PM_{2.5}$ ratio	1.46	1.63	1.77	1.88	2.02	2.14	2.28	2.58	3.03	4.01	5.72
PM_{coarse} ratio	1.68	2.05	2.32	2.73	3.24	3.82	4.42	5.54	8.01	14.34	52.52

2.2 Population Coefficient of Variation

This parameter measures the amount of random, day-to-day movement of the true concentration about the average sine curve. The population coefficient of variation (CV) parameter must be estimated separately for the $PM_{2.5}$ and PM_{coarse} series. This parameter is a bit harder to estimate than the ratio parameter. The following procedure does a reasonable job of estimating the parameter. Starting with every 6th day measurements (deleting, if necessary), the natural log of each concentration is taken. Next, a new sequence of numbers is generated with the differences of successive pairs in the sequence of the logs. Every other term in this sequence is removed so that each term is independent of the others. Let *S6* = the standard deviation of this set of numbers. An estimate for the population CV is $\sqrt{\left(\exp\left(S6^2/2\right)-1\right)}$. Table 2 shows quantiles of the estimated population CV parameter for both $PM_{2.5}$ and PM_{coarse} across several sites in the United States.

Table 2. Population CV Quantiles

Quantile	2.5	10	20	30	40	50	60	70	80	90	97.5
$PM_{2.5}$ CV	0.35	0.41	0.45	0.48	0.51	0.53	0.56	0.6	0.64	0.69	0.8
PM_{coarse} CV	0.4	0.49	0.56	0.61	0.66	0.71	0.76	0.84	0.93	1.08	1.39

2.3 Autocorrelation

Another parameter describing the natural variability of the true concentrations is autocorrelation. Like the preceding variables, the autocorrelation must be estimated separately for the $PM_{2.5}$ and PM_{coarse} series. This is a measurement of the similarity between successive days. Estimating autocorrelation is even harder than estimating the population CV. It is only estimated from daily measurements. Otherwise, 0 can be used as a conservative case. For the sites that did have daily measurements, let *S*6 be the standard deviation computed as in the section on population CV based on differences of the logs from every 6th day measurements. Let *S*1 be the corresponding standard deviation calculated using differences of logs from daily measurements. If *S*6 > *S*1, the autocorrelation is estimated by $\left(S6^2 - S1^2\right)/S6^2$, otherwise it is estimated by 0. (This estimation tends to slightly overestimate the truth. Since it is better to underestimate this parameter (to make the results more conservative), future estimates may be multiplied by 0.85. Note that the usual autocorrelation estimate, correlation estimate between successive values, does not work when there is seasonality.) Table 3 shows quantiles of the estimated autocorrelation for both $PM_{2.5}$ and PM_{coarse} across the 65 sites in the database with daily measurements.

Table 3. Autocorrelation Quantiles

Quantile	2.5	10	20	30	40	50	60	70	80	90	97.5
$PM_{2.5}$ autocorrelation	0	0.05	0.21	0.3	0.36	0.38	0.41	0.44	0.51	0.58	0.8
PM_{coarse} autocorrelation	0	0.11	0.16	0.19	0.24	0.27	0.38	0.4	0.43	0.54	0.69

2.4 PM_{coarse} to $PM_{2.5}$ Ratio

This parameter is used for scaling. Its estimation is very simple. Let $M1$ be the average of all of the PM_{coarse} values over the full time period available. Let $M2$ be the average for the corresponding $PM_{2.5}$ data. Then, k is estimated with $M1/M2$. Table 4 shows quantiles of the estimated value of this multiplicative factor across the sites in the database.

Table 4. PM_{coarse} to $PM_{2.5}$ Ratio Quantiles

Quantile	2.5	10	20	30	40	50	60	70	80	90	97.5
k	0.28	0.37	0.46	0.56	0.72	0.87	1.04	1.29	1.6	2.22	3.29

2.5 Offset Between $PM_{2.5}$ and PM_{coarse} Cycles

This parameter controls the difference in time between the peak of $PM_{2.5}$ in a year and the peak of PM_{coarse} (and, by implication, PM_{10}) in a year. The value of the parameter is in months. In order to estimate this parameter, monthly means are calculated as in the estimate for the seasonality ratio parameter. The offset parameter estimate is the number of the month with the highest average PM_{coarse} level minus the number of the month with the highest average $PM_{2.5}$ level. Since the sine wave is cyclical, adding or subtracting 12 from the estimate does not change the sine wave produced. For consistency, all of the estimates were converted to equivalent values between -5 and 6 in Figure 2.

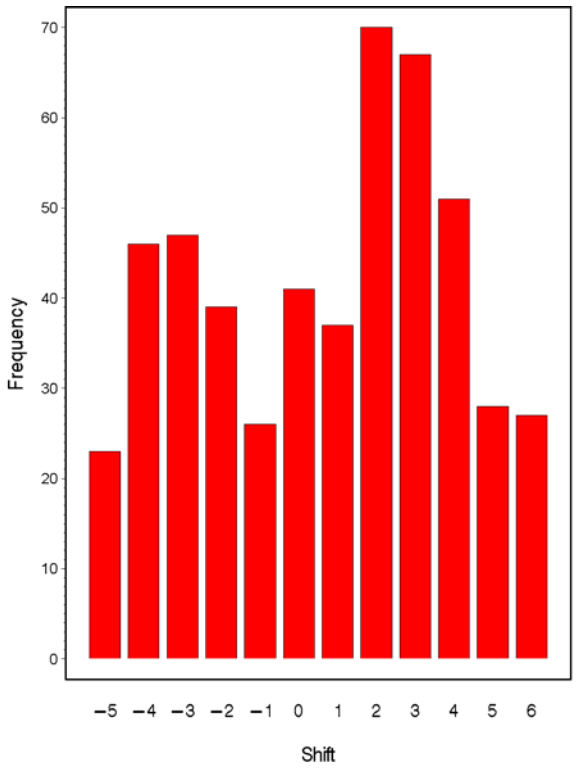

Figure 2: Bar Chart of the Offset Parameter.

2.6 Correlation between $PM_{2.5}$ and PM_{coarse}

This parameter estimates the correlation between the two series. The calculation of this correlation can be affected by autocorrelation and seasonality in the series, so the calculation is quite complex. First, consider only the $PM_{2.5}$ series. Starting with every 6th day measurements (deleting, if necessary), the natural log of each concentration is taken. Next, a new sequence of numbers is generated by the differences of successive pairs in the sequence of the logs. Every other term is then removed from the sequence. This procedure is repeated with the PM_{coarse} concentration sequence. Then, the corresponding elements from these two sequences are added together to obtain a single sequence. Let *SS* be the standard deviation of this set of numbers. Next, obtain the standard deviation values called *S*6 in the population CV section for the $PM_{2.5}$ and PM_{coarse} sequences. Call these $S6_{25}$ and $S6_{coarse}$ for the $PM_{2.5}$ and PM_{coarse} sequences, respectively. The correlation can be estimated by $[SS^2 - (S6_{coarse}^2 + S6_{25}^2)] \Big/ (2 * \sqrt{S6_{coarse}^2 \times S6_{25}^2})$. (The usual correlation estimate between individual PM measurements from the two series will overestimate correlation in the presence of autocorrelation and seasonality.) Table 5 shows quantiles of the estimated value of the autocorrelation across the sites in the database.

Table 5. PM_{coarse} to $PM_{2.5}$ Correlation Quantiles

Quantile	2.5	10	20	30	40	50	60	70	80	90	97.5
correlation	-0.23	-0.05	0.06	0.12	0.19	0.25	0.31	0.39	0.46	0.56	0.69

2.7 Type I and Type II Errors

Type I and Type II errors describe the probability of making the wrong decision under a specified set of conditions. Type I error is the probability of observing a (three-year aggregate) value above the standard when the true ambient level (free from measurement error and bias) is below the standard. In the tool, this parameter is required to be at least 1 percent; otherwise more simulations are needed to get robust results. (See Step 6 of U.S. EPA's QA/G-4 for additional guidance.)

Type II error is the probability of making the opposite mistake: observing a value below the action limit when the true ambient level (free from measurement error and bias) is above the action limit. Since the curves show the probability of observing a value above the action limit, the value of 1 minus the Type II error is shown at the top of the graphs. As with the Type I error, this parameter is required to be at least 1 percent; otherwise more simulations are needed to get robust results. (See Step 6 of U.S. EPA's QA/G-4 for additional guidance.)

2.8 Standards

Both a daily standard and an annual standard can be entered for PM_{coarse}. The DQO tool is intended to help answer questions of the type, "What is the probability of the observed 98th percentile of PM_{coarse} levels exceeding 40 $\mu g/m^3$ when the true 98th percentile of PM_{coarse} levels is 37 $\mu g/m^3$?" In this question, 40 $\mu g/m^3$ is the daily standard and 0.98 is the percentile for the daily standard.

2.9 Sampling Frequency

The sampling frequency is entered in terms of period between sampling events. The value of *m* must be an integer from 1 to 12 and denotes the number of days between successive samples. The most common values are 1, 3, 6, and 12. A value of *m* equal to 6 corresponds to approximately 15 sampling days each quarter.

2.10 Completeness

Completeness is the minimum acceptable percentage of the data that is intended to be collected. Completeness is included in the DQO tool to mimic random occurrences of data loss, such as a power outage on a scheduled sampling day. The criterion is applied quarterly. Thus, if the completeness is set to 0.75, the DQO tool removes 75 percent of your data from each quarter of each year. However, the completeness requirements are on the $PM_{2.5}$ and the PM_{10} measurements. Hence, there is the possibility (in fact, it is almost always the case in the simulations) that the percent of the PM_{coarse} measurements used is less than 75 percent since different days may be missing in the two sets of measurements.

2.11 Bias

The bias input is the maximum allowable absolute measurement bias as a proportion of the truth for $PM_{2.5}$ and PM_{10}. Bias is a consistent measurement error — a tendency to always either overestimate or underestimate the truth. The DQO tool accepts only positive values for quantifying bias. Both positive and negative biases are simulated.

2.12 Measurement Coefficient of Variation

Measurement CV quantifies the size of the measurement error. It is expressed as a proportion of truth for $PM_{2.5}$ and PM_{10}. The random component to the measurement error is assumed to follow a normal distribution with a mean of 0 and a standard deviation that is proportional to truth (for the given day).

3.0 References

U.S. EPA (2000). "Guidance for the Data Quality Objectives Process (EPA QA/G-4)." Report No. EPA/600/R-96/055, August.

U.S. EPA (2002). "DQO *Companion*, Version 1.0 User's Guide," written by Battelle for U.S. EPA under Contract No. 68-D-98-030, Work Assignment 5-07.

The QA Strategy Workgroup's Validation Templates

Tom Parsons, South Coast Air Quality Management District

The National QA Working Group has been given the task of updating and clarifying the Quality Assurance requirements for ambient air monitoring found in 40 CFR 58 Appendix A. The approach the working group has taken is to develop comprehensive data validation templates for all the criteria pollutants similar to the template released for the $PM_{2.5}$ program. This presentation will cover the rationale behind and the organization of the templates, the current version of the template for each pollutant, and then provide an opportunity for members of the audience to make suggestions for improvements to the templates.

Driving Excellence - Quality Methods at the National Vehicle and Fuels Emissions Laboratory

Thomas Schrodt
Manager, Compliance and Development Testing
USEPA - National Vehicle and Fuels Emissions Laboratory

Americans love their vehicles and drive more and more each year. Creating change that lessens the environmental impact of this huge sector is a high stakes endeavor from both an environmental and economic perspective. Information quality is vital to this effort. A multi-faceted approach is taken for Quality Control of vehicle testing operations at the EPA National Vehicle and Fuels Emissions Laboratory, in Ann Arbor, Michigan. Major components of the program include Process Development and Review, Standard Operating Procedures, Calibration Assurance, Change Management, Test Packet Audits, System Audits, Laboratory Correlation, Diagnostic and Statistical Tools, and Continuous Improvement. The inherent test to test variability of vehicles themselves present specific challenges and require special consideration to stratify sources of variation in the measurement system. Quality expertise at the laboratory is leveraged through various technical exchange and standards committees, Cooperative Research and Development Agreement activities and external auditing. Future challenges are driven by increasing regulatory stringency and complexity.

Background

The U.S. Environmental Protection Agency's (EPA) Office of Transportation and Air Quality's (OTAQ) mission is to reconcile the transportation sector with the environment by advancing clean fuels and technology, and working to promote more liveable communities. OTAQ is responsible for carrying out laws to control air pollution from motor vehicles, engines, and their fuels. Mobile sources include: cars and light trucks, large trucks and buses, farm and construction equipment, lawn and garden equipment, marine engines, aircraft, and locomotives.

EPA's motor vehicle emissions control program was established in 1971. OTAQ staff is divided between EPA's headquarters in Washington, D.C., and the National Vehicle and Fuel Emissions Laboratory (NVFEL) in Ann Arbor, Michigan, near the headquarters of domestic automobile manufacturers.

There are about 400 employees in OTAQ. Staff expertise spans a variety of technical and public policy fields including auto mechanics, engineering, chemistry, economics, natural resources management, and law. OTAQ develops national standards for emissions, evaluates emission control technology, tests vehicles, engines, and fuels, and determines compliance with federal emission and fuel economy standards. We also develop fuel standards, guidance for state inspection and maintenance programs, and market and transportation incentive programs.

The early goals of OTAQ centered around working with industry to reduce emissions from individual automobiles. The approach achieved dramatic success-compared to an uncontrolled

passenger car of 1970, an average car on the road today emits 60 to 90 percent less pollution over its lifetime. However, the amount of driving in this country has more than doubled since then, so transportation still accounts for a large part of national air pollution. About half the U.S. population live in areas where pollution levels exceed federal air quality standards.

The National Vehicle and Fuel Emissions Laboratory in Ann Arbor, Michigan had it's beginnings before the formation of the Environmental Protection Agency. During this 30+ year period, creativity innovation and quality have been hallmarks of the many test programs conducted in the areas of compliance monitoring and enforcement, regulatory development, energy efficiency, technology demonstration and the assessment of innovative approaches to motor vehicle emissions measurement and control. The work of the laboratory has been key to the enormous strides made to clean the air through technological developments to motor vehicles and their fuels. The purpose of this paper is to describe the work of the laboratory through presentation of several key programs along with various levels of Quality Assurance and Quality Control that ensure the veracity and effectiveness of these programs.

The laboratory facility consists of about 160,000 square feet of office space and laboratory functions including chassis and engine dynamometers and a comprehensive fuels laboratory. Increased regulatory stringency and the emerging technological sophistication of engines and vehicle systems have placed increasing demands on the science of emissions and performance measurement.

Historically the emphasis of testing has been light duty vehicles, but in the last ten years engine-based programs have grown extensively and now comprise a significant portion of the testing effort. This shift in emphasis is both the product of a wider regulatory efforts and increased focus on increasing the efficiency of the internal combustion engine to help to control the emission of greenhouse gases. Engine testing ranges from 1 horsepower "weed whacker" engines to heavy duty diesel engines of greater than 300 horsepower, covering both "on-highway" applications and an expanding class of "non-road" applications.

One of the longest standing programs, and historically a central emphasis of the laboratory, is the pre-production certification of new vehicles for compliance with emissions and fuel economy standards. This process is largely a self certification process done by the auto manufacturers, but the laboratory plays an important role by conducting confirmatory testing of a random selection of certification vehicles. This role is essential, due the lack of a single reference standard for the measurement of automotive emissions. As part of this program the laboratory tests prototype vehicles to assure that they meet the appropriate regulatory limits and that the comparison of EPA and manufacturer results indicate no long term bias. As such the laboratory serves as the national performance standard for automotive emissions and fuel economy testing.

A program which is closely related, and actually reinforces the pre-production certification program, is the surveillance testing of in-use vehicles. Automobile manufacturers are responsible for the emissions performance of their vehicle for the full useful life of the vehicle or 100,000 miles. At NVFEL, random selections of available vehicles within targeted engine families are made and these vehicles are tested to ascertain emission levels. If these vehicles shows evidence of pattern-type failure, a larger selection of this engine family is made to confirm

the results. In a typical year this testing inspires the voluntary recall of about 1 million vehicles. It is very rare that EPA has to take any sort of legal action to motivate these recalls, because of the known quality of our program and laboratory data.

NVFEL also conducts various types of testing of in-use vehicles to help build computer models of emissions performance of vehicle populations to assist states and local governments in making choices as to the most effective and efficient approaches to reducing air pollution in their localities.

EPA-OTAQ is in the midst of developing and implementing several new regulations that reflect the great strides being made to reduce emissions from internal combustion engines. The passage of these regulations has relied heavily on the demonstration of new technology at NVFEL. When EPA was sued regarding the feasibility of 2007 diesel engine standards, the court relied heavily on NVFEL test results in upholding the regulation. Similarly when the EPA developed the Tier 2 regulation, which is dramatically reducing emissions from light and medium duty vehicles, it was the development and demonstration of successful control strategies at NVFEL which convinced automobile manufacturers of the feasability and effectiveness of the controls. In the heavy duty arena, incredible reductions are being made in the control of both particulate matter and oxides of nitrogen from diesel engines through NVFEL technology demonstrations.

Leveraging it's role as the performance standard for automotive emissions and fuel economy testing, EPA is also at the forefront of cooperative activities with industry aimed at improvements to measurement processes through both informal and formal mechanisms such as Cooperative Research and Development Agreements (CRADAs) and formal standards committees such as those sponsored by the Society of Automotive Engineers.

NVFEL Quality Methodologies for Laboratory Operations

Quality tools for the management and continuous improvement of NVFEL laboratory operations have been developed and refined over many years. Some of these tools are very similar to those found in most laboratory environments, others have been shaped by some of the more unique conditions present in the testing of automotive emissions and fuel economy. The following is a review of some of the major on-going elements that form a basis of our laboratory quality control activities.

Process Development, Review, and Standard Operating Procedures

New measurement processes and systems undergo review by a team of laboratory experts to assure that the new processes will achieve the desired results. There are many significant and specific requirements in associated regulations that form the basis for part of this review. A significant, ultimate outcome of this review is to translate regulations into detailed procedures and data forms. The resulting procedures are not only used internally, they are available to anyone on the internet at http://www.epa.gov/otaq/testproc.htm., and are used by various automotive emissions testing organizations as a basis for their own forms, procedures and quality methods.

Calibration Assurance

Calibration, and care to maintain traceability, are major issues for most laboratories. To assure a consistent approach across all laboratory operations and to track routine activities required to assure proper equipment calibration, we developed our own equipment tracking system. This system automatically generates work orders for calibration efforts based on pre-determined schedules and needs and provides a management reporting system for calibration and maintenance. This ensures both the timely calibration of equipment, and the maintenance of traceability. This system has proven itself so valuable that it has been successfully adapted to track test equipment preventative maintenance, and inspection and maintenance of about 100 safety-related items in the lab as well.

Change Management

To manage change and systematize the results of continuous improvement in the laboratory, NVFEL developed and implemented a system known as the Equipment and Procedure Change Notice system (EPCN) which ties directly in to our new process review efforts. This system implements and guides various changes. For example the EPCN for a new analytical system will spell out all the necessary actions needed to implement that new system. All the various testing, process development efforts and quality control efforts related to the new site are driven by the EPCN. This process is administered by our Quality Control Staff.

Test Packet Audits

On a random basis, all official test result documents are audited to ensure that proper procedures and forms have been used for a particular kind of test. These audits are highly structured to cover all CFR and internal requirements. These periodic audits catch any emerging trends, which could indicate that systematic mistakes are made and also serve as a driver for continuous improvement efforts. Additionally all test packets for testing done to confirm suspected in-use vehicle pattern failures are audited to ensure that all official documentation is bullet proof.

System Audits

System audits are targeted to take an in-depth look at specific areas such as qualification of new measurement systems or the complete review of all processes involved in recall testing. External audits and any resulting resolution activities are included in this category. Historically the laboratory portion of the organization has been judged to have a robust quality system that has formed an underpinning of the work of the other divisions within OTAQ.

Laboratory Correlation Activities

Two conditions are present in automotive emissions testing that make the use of structured correlation programs vital to an on-going assessment of measurement quality. First since the measurement is of a complex operating system, an automobile, it is very difficult to segregate measurement variability from the variability of the item being tested. Second since there is a

highly complex interaction of the measurement system and the item being measured, no single physical standard exists for the outcome of the test. It is the combination of these factors that makes the performance-standard role of NVFEL so important.

To help carry out the role of performance standard, the laboratory relies on three levels of correlation testing. At the first level the laboratory uses a highly developed, statistically based program of intra-laboratory correlation testing using a specially selected and instrumented vehicle. This vehicle is run weekly on all active chassis dynamometers to look for immediate offsets from normal results, trends or other unusual changes. This is a very important aspect of the laboratory's system of diagnostic and quality control system because it exercises all components of the measurement system.

The laboratory also continuously monitors the correlation of specific vehicle or engine test results with other comparable facilities. This is primary done through two mechanisms, "Paired Data" and structured inter-laboratory correlation programs.

Paired Data compares EPA confirmation test results against comparable tests generated by the manufacturer with the same vehicle. Offsets and trends are monitored, and the manufacturers are subject to increased levels of confirmatory testing of all their certification vehicles if their data are persistently different. This process also assures EPA that our measurement systems are functioning properly through continuous comparison with a sizeable population of laboratories.

Structured inter-laboratory correlation testing uses dedicated vehicles or engines to compare results between EPA and other laboratories, in a manner which is structured to identify sources and magnitudes of variation. This testing is particularly important for new or modified test procedures and for the more stringent requirements in recent EPA regulations. Wherever possible, EPA strives to participate with groups of manufacturers rather than individual entities to improve the quality and efficiency of the program. This program also serves as a back drop to a whole range of cooperative efforts aimed at exchange of information and continuous improvement of measurement processes.

Diagnostic and Statistical Tools

Regulatory requirements, calibration requirements, documentation, and procedures are critical to the quality and legality of certain kinds of testing, but they are not always sufficient or efficient in terms of achieving test quality. NVFEL performs specialized diagnostic tests which can be run more frequently and efficiently than calibrations and which exercise entire parts of the test system thus checking equipment, calibration, and operation in one pass. These tests are performed, and assessed regularly by Quality Control staff and a variety of engineers and technicians. These tests provide confidence in the readiness of equipment, act as an early warning signal and halt testing if the results are not within expectations, and provide a basis for adjusting calibration intervals and requirements if that is appropriate. To facilitate the analysis of diagnostic test results a variety of statistical tests and methods are used including Analysis of Variance and control charting.

Continuous Improvement and Future Challenges

Quality of organizational outcomes is not static. There is no status quo. The organization is either improving the quality of it's outcomes or, things are getting worse. Efforts to make things better at NVFEL center both on continuous improvement of existing processes and products, and applying new approaches to meet the challenges of measuring the lower emissions that we have helped to make possible. Some examples of improvement projects now underway are:

- Gaining efficiency and effectiveness by bringing together common or similar elements of quality, safety and environmental management systems
- Improving program and project planning to maximize outcome success
- Management of gas analyzer calibration and maintenance of gas standards at very low levels
- Improved emission sampling systems and processes

Quality Assurance or Quality Assistance?

Louis Blume, U.S. Environmental Protection Agency, Great Lakes National Program Office, 77 West Jackson Boulevard, Chicago, Illinois 60604 and Judy Schofield, DynCorp, Science & Engineering Group, 6101 Stevenson Avenue, Alexandria, VA 22304

Too often researchers or decision makers are asked to develop a Quality Management Plan and believe that the process of doing this is just another requirement to "get the money" or to make a bureaucrat happy. These feelings typically are driven by past experiences of working with quality documentation or quality managers that were inflexible and rigid in their interpretation of requirements with the hopes of guaranteeing or assuring success in a project, similar to an insurance policy or a guarantee.

The authors believe that the Agency's current Quality System developed within the constraints of the ISO 9000:2000 and ISO 9001: Quality Management Systems-Requirements provides for a time saving cost effective planning and implementation process that adds value to the decisions or questions being addressed. Further the Quality staff employed within a successful organization provide a critical assistance role to the researcher or environmental manager. As an organization matures in applying these successful quality systems in a flexible manner that focuses on assistance, the historic polarization between the client and the quality assistors is greatly reduced. This promotes continued interaction between the client and quality assistors well beyond the initial award phase of the project to the reporting, assessment, and peer review phases. This continued interaction helps in assuring that results are consistent with the requirements of OMB's new Information Quality Guidelines.

An Intra-active Web Based Research Plan/Quality Assurance Plan

Allan R. Batterman, Quality Assurance Manager, NHEERL, MED

A web based Research Plan/Quality Assurance Project Plan (RP/QAPP) document is interactively linked to documentation that asks pertinent questions and gives suggested examples and formats, to assist Research Scientists in developing a RP/QAPP of correct detail and format to facilitate rapid approval so they can begin their research unimpeded.

This documentation is listed on the MED QA Webpage; the document will be demonstrated in this presentation.

Introduction

At the Mid-Continent Ecology Division, the research staff has always struggled with the writing of the QAPP, primarily because several parts of the separate QAPP are duplicative of the Research Plan and the specific information being asked for in the QAPP is somewhat foreign to the researcher. Historically we have allowed researchers to simply reference in the QAPP that section of the Research Plan, i.e. "Experimental Design," however this has meant that the reviewers need to read both documents throughly to get a full comprehension of the project. This is a step wasteful of reviewer's time. Over the last several years, I have tried various processes to limit the revisions required to get the necessary information and detail into the QAPP. The best method seemed to be combining the two documents and providing guidance directly, which is relatively easy to accomplish, with Hyper Links to gain access to Guidance Documentation (G Series and others) and Frequently Asked Questions (FAQs). During this presentation this package will be explained and demonstrated.

Background

ORD has incorporated into its research program the various Research Strategy Documents, which explain the basic ORD research program areas. The Research Scientist is then required to develop the Research Plan/Quality Assurance Program Plan that will cover their work. At MED, we have found that often RP and QAPP need to be revised several times to assure inclusion of all required materials. By combining the RP and QAPP into one document we have lessened the duplication of materials. Then, by Hyper Linking guidance and FAQs to the various associated sections, the researcher can see exactly what is to be discussed and considered as they write those sections. The reference materials and FAQs are easily available, at a click of a computer mouse, while writing in the pre- formatted document. The document is then programmed in such a manner that these links disappear when it is saved. To accomplish this, a complete understanding of the process was required, as well as a thorough understanding of the guidance and aids that are available to assist the writer. This document is constantly modified as new and revised documentation is continually being developed. The RP/QAPP also provides examples of accepted packages by QA Category Level to show how others address the concerns. These examples are provided in an edited manner that does not simply provide the writer with a boiler

plate to incorporate.

This document is available on the MED Intranet Web page under the Research Plan button and is accessible from the NHEERL QA Intranet Page at the MED Link.

Planning for Quality in EPA's Next Generation Mobile Source Emissions Model (MOVES)

Chad Bailey, Dave Brzezinski, Mitch Cumberworth,
Constance Hart, John Koupal, and Harvey Michaels
U.S. EPA Office of Transportation and Air Quality

Introduction

The Clean Air Act requires EPA to develop and regularly update emission factors for all emission sources. Pursuant to this charge, EPA's Office of Transportation and Air Quality (OTAQ) has developed a number of emission and emission factor estimation tools for mobile sources, including MOBILE (for highway vehicles) and NONROAD (for off-road mobile source pollutants). EPA is proposing to update these tools with the Multi-Scale Motor Vehicle and Equipment Emission System (MOVES), which is intended to include and improve upon the capability of these tools and, eventually, to replace them with a single, comprehensive modeling system.

The National Research Council (NRC) published a thorough review of EPA's mobile source modeling program in 2000 (NRC, 2000).[i] The NRC provided several recommendations for improving EPA's mobile source modeling tools, including (a) the development of a modeling system more capable of supporting smaller-scale analyses; (b) improved characterization of emissions from high-emitting vehicles, heavy-duty vehicles, and off-road sources; (c) improved characterization of particulate matter and toxic emissions; (d) improved model evaluation and uncertainty assessments; and (e) a long-term planning effort coordinated with other governmental entities engaged in emissions modeling.

A particular focus of the NRC report was the need to provide emission factors and tools that will allow the estimation of emissions at finer analysis scales. Historically, EPA's mobile source emission estimation tools and underlying emission factors have been focused on the estimation of mobile source emissions based on average operating characteristics over broad geographical areas. Examples of this scale of analysis are the development of SIP inventories for urban nonattainment areas and the estimation of nationwide emissions to assess overall trends.

In recent years, however, analysis needs have expanded in response to statutory requirements that demand the development of finer-scale modeling approaches to support more localized emission assessments. Examples include "hot-spot" analyses for transportation conformity and evaluation of the impact of specific changes in transportation systems (e.g., signalization and lane additions) on emissions.

We have adopted most of the NRC's recommendations as our objectives in designing MOVES. The MOVES design objectives include applicability to a wide range of spatial and temporal scales, inclusion of all mobile sources and all pollutants, addressing model validation and the calculation of uncertainty, ease of updating the model, quality assurance, ability to interface with other models, and ease of use.

These objectives have shaped our development plan for MOVES, as detailed in the rest of this paper.

We have begun and will continue to develop MOVES with extensive coordination and outreach. A cross-agency team representing OTAQ, the Office of Air Quality Planning and Standards (OAQPS), the Office of Research and Development (ORD), and Regional EPA offices

produced an issue paper containing an initial proposal for the model in April 2001.2 Since then, we have coordinated with states, metropolitan planning organizations, the U.S. Department of Transportation, consultants, academics, and the "FACA Modeling Workgroup", which is the Modeling Workgroup of the Mobile Source Technical Review Subcommittee (MSTRS) of the Clean Air Act Advisory Committee (CAAAC), a committee of experts from government, industry, and academia established under the Federal Advisory Committee Act (FACA). This continuing coordination will enable MOVES to benefit from advances in mobile source emission modeling and to meet the needs of the user and stakeholder community. In addition, EPA has funded independent external peer review of the MOVES model via a contract with Southwest Research Institute (SwRI). SwRI selected panelists from outside the EPA with expertise in transportation planning, emissions modeling, and air quality analysis. The panel has provided comments on software design and emission data analysis plans (EPA, 2002a; EPA, 2002b).[ii],[iii]

This paper provides an overview of the MOVES development process with particular eye toward model quality (EPA, 2002). First, the intended requirements and applications of the model are discussed. Following, is a discussion of general model design considerations, its theory in emissions research, input data indicators of data quality and completeness, data acquisition and management in the Mobile Source Observation Database (MSOD), input quality indicators included in MOVES, and considerations of future data collection processes.

Use Cases for MOVES

As the first part of model design and quality planning, a complete understanding of requirements for the model is needed. OTAQ prepared a list of "use cases," a definition of the multiple ways in which mobile source emissions models are used for regulatory and research purposes. With the help of a contractor (MCNC), we interviewed expert users for each use case. In addition to this formal interviewing process, we also consulted closely with researchers and users within EPA's Office or Research and Development (ORD) and Office of Air Quality Planning and Standards (OAQPS), as well as with the FACA Modeling Workgroup, comprised of experts from industry, academia, government, and consulting firms. From this assessment we developed a list of "essential use cases," which explore, at a high level, how users need to or would like to use the model. The result of this analysis was a list of fundamental use cases that have driven the MOVES design and quality assurance processes. These use cases include national inventory development for EPA reports and regulation, local inventory development (e.g. State Implementation Plans), hot-spot and project level analysis (e.g. National Environmental Policy Act), interaction with air quality and travel models, policy evaluation, model analysis (e.g. uncertainty analysis), and model update and expansion.[iv]

As discussed below, these considerations are the first step in model design. Crucially, these use cases also assist in defining the decisions that rely upon EPA's mobile source emissions models. These decisions are key to defining the requirements for MOVES quality assurance process. As noted in the April 2002 peer review draft of Guidance on Quality Assurance Project Plans for Modeling,

> *The intended use of the model is a determining factor in the level of QA needed because it is an indication of the seriousness of the potential consequences or impacts that might occur due to quality problems… [other] aspects of the QA effort can be established by considering the scope and magnitude of the project.*

The use case definitions also help in establishing the rigor of the QA process. Mobile source emissions models play a central role in a wide variety of policy decisions and analyses, including

development of inventories for state implementation plans (SIPs), the National Emissions Inventory (NEI) used in the Trends report, conformity decisions that affect transportation funding, and project-level analyses undertaken through the National Environmental Policy Act. As a result, quality assurance planning for MOVES has been an integral part of model planning and development. This process began by convening meetings of each of the MOVES teams, the Design Team, the Software Implementation Team, and the Mobile Emissions Analysis Team (MEAT) to determine which of the requirements outlined in the draft QAPP guidance for modeling were the responsibility of each model team. Each of these teams has developed its own QA processes, which are now being assembled in a QAPP for the entire model.

Model Design

The use cases presented in Table 1 reflect overarching design considerations in the development of MOVES. MOVES will be developed on an iterative basis, with most important use cases addressed first.

The overall design of MOVES is modular, general purpose, "data-driven", easy-to-use, and high-performance. All geographic scales and processes are incorporated into a flexible framework of time spans and locations. The MOVES design provides several means of modeling the effects of emission controls. MOVES emission rates and activity information are derived from databases, easily updated as needed. The first part of model description is to define what the model is intended to do. Fundamentally, MOVES predicts emissions of all pollutants from multiple emission properties from all mobile sources addressed by MOBILE and NONROAD at a range of geographic scales for a range of durations. In order to do this, MOVES applies the basic concept that for a given time, location, use type, and emission process, total emissions can be calculated in four steps: 1) calculate total activity, 2) determine activity distribution within source and operating mode bins, 3) calculate an emission rate for each process, source bin, and operating mode, and 4) aggregate emission rates across these modes.

The term *source* is used to encompass both on-road vehicles and off-road equipment pieces. A *source use type* is a specific class of on-road vehicles or off-road equipment defined by unique activity patterns. *Source bins* are a subset of use types: subcategories that differentiate emission levels within a use type, covering categorizations such as weight class, fuel type, technology, standard, horsepower range, etc. *Total activity* is defined for a given use type as the product of the population of that use type and the per-source activity by time and location. We subdivide total activity into categories that differentiate emissions, known as operating mode bins; the intersection of source bins and operating mode bins results in a unique source and operating mode bin. By *emission rates*, we mean the most disaggregated rates the model produces internally by source and operating mode bins. By *emission factors*, we mean emission rates aggregated in various ways over source and/or operating mode bin and normalized by some activity basis, such as mass of pollutant per time or per mile. An *emission process* is a unique emissions pathway. Generally speaking, emissions for a given emission process may be calculated by multiplying activity in each source and operating mode bin by the emission rates for that bin and aggregating upward.

Fundamental to MOVES is the transition into a relational database framework for emissions estimation. Built into the model is a series of databases, including a database containing emission rates for each bin. This design allows for the emission rate databases to be

updated without needing to change the underlying code of the model.

Model Theory

For this paper, a full description of MOVES' overall theory is unnecessary. However, the process by which OTAQ decided upon a particular design and theoretical framework are important in illustrating the attention to quality concerns.

In July 2001, OTAQ posted a RFQ to analyze real-time emissions data (1-Hz frequency) collected as part of an effort to develop portable on-board emissions measurement devices. EPA awarded contracts to the University of California-Riverside, ENVIRON, and North Carolina State University. The awardees were provided with on-board emissions, activity, and location data from several hours of data collected on 12 light-duty gasoline vehicles, heavy-duty diesel vehicles, and off-road diesel engines. The awardees proposed analytical methods and submitted their predictions for a set of "emissions-blinded" data to EPA. Proposals were evaluated according to a set of criteria, including accuracy of the method (verified at a "proof of concept" level), applicability to MOVES, extensibility to the range of vehicles and fuels likely in current and future technologies, usability of a wide range of data sources including laboratory and on-board data, and ease of update as new data becomes available.

On the basis of this "shootout," NCSU was selected to receive a second contract to analyze a different set of emissions data, including continuous on-board, dynamometer, remote sensing device (RSD), and IM data. The project evaluated methods for developing modal emission rates from these data sources for a data set of light duty gasoline vehicles. NCSU was also charged with suggesting a methodology for quantifying both variability and uncertainty in emissions. The methods were calibrated on a small data set and compared with a subset of other vehicles. NCSU then developed a recommendation for developing modal emission rates based on vehicle specific power (VSP) for use in MOVES.[v] In this framework, total emissions could be calculated by multiplying estimated time spent in each VSP bin (activity) by the bin-specific emission rate and adding across all source and VSP bins. NCSU proposed that variability in each bin could be characterized by fitting parametric distributions to the data within each bin. Uncertainty can be characterized using standard error propagation methods.

As an alternative, EPA also has developed a plan to rely on a physical emissions model based in part on the Comprehensive Modal Emissions Model (CMEM) developed by UC-Riverside.[vi] PERE will provide emission rate estimates in source and operating mode bins in which insufficient measurement data is available, such as is the case for future vehicle technologies. As proposed, the Physical Emission Rate Estimator (PERE) would 1) predict fuel rate based on vehicle power and engine friction; 2) predict of engine-out emissions from fuel rate; and 3) estimate aftertreatment effectiveness (e.g. catalytic efficiency) from fuel rate. The relationships in these three steps would be based on parametric equations, and the model would produce binned emission rates by applying these equations at the midpoints of the appropriate operating mode bins. Methods for characterizing uncertainty in PERE may include traditional error propagation.

Quality Indicators for MOVES Input Data

Critical to MOVES is the availability of emissions data that is well documented and complete in its coverage of variables of interest. As a result, in collection of all new data for

inclusion in MOVES, a series of "quality and completeness" criteria have been developed to categorize each source of data. The preponderance of available data for MOVES has been gathered by organizations outside of OTAQ. As a result, it is necessary to determine whether measurement projects were well documented, and whether the data produced contains information deemed necessary for MOVES in estimating emissions. As part of MOVES data collection, data are assigned a letter grade of A, B, or C, one grade for availability of complete documentation of the data source and another grade for completeness of the data reported to EPA.

Documentation grades are assigned based on the availability of documentation of the data collection process from each data source. For each test program, OTAQ wants copies of the statement of work, QA/QC plans, program reports summarizing the results of the test program and any changes made to the initial test plan, descriptions of instrumentation, and documentation of measurement uncertainty, including instrument minimum detection limits and reproducibility of data. An "A" grade for documentation is assigned to the data when all desired documentation exists and is available upon request. A "B" grade is assigned when all desired documentation can be derived from testing records and charts. A "C" grade is assigned when some of the desired documentation is missing.

Data received completeness scores based on whether the test program in they were obtained recorded all variables determined to be critical needs for MOVES, including information on vehicle or engine characteristics, pollutants measured, fuel parameters, and descriptions of the testing conditions (e.g. soak time before engine start). Data receive an "A" if all critical fields are measured and available, a "B" if all critical fields are either measured or can be derived from reported data, and a "C" if some critical fields were not measured and cannot be derived from reported data.

These documentation and completeness scores allow MOVES data to be categorized with regard to suitability for inclusion in the model. The MOVES emission analysis team is currently developing acceptance criteria for how each grade of data will be treated (or not treated) within the model.

Data Storage and Management

OTAQ is home to the Mobile Source Observation Database (MSOD), a centralized repository of any type of vehicle measurement data.[vii] Developed since 1998, the MSOD is a relational database that is paired with a formal data acquisition and management process. MOVES will use MSOD as its storage facility for all emissions data.

MSOD is a relational database that has been implemented in Oracle to store emissions data on an aggregate (driving cycle average) basis or on a continuous basis, which makes it suitable for MOVES. MSOD. MSOD currently stores data on observations of in-use mobile sources, including general-purpose emission factor data collected since 1982 as well as data from special studies. The general design allows storage of any type of mobile source observation, including emission rate data, activity, and fuel parameters. MSOD uses a unique mobile source identification number for each vehicle and emission test that maps to specific contracts and statements of work. Stored in the database are work assignments numbers, instrumentation used during measurement, and contract numbers. Repeated observations of the same variable are stored to allow for repeatability to be assessed.

The MSOD quality assurance system is well established and broad. Its data acquisition and management process allows for identification of new research needs (as identified by

management or MOVES data needs) and translation of these needs into short-term and long-term objectives and data needs. These needs are incorporated into the annual and long-term information acquisition planning, which involves prioritizing needs, determining the appropriate technical approach, developing a data management plan for each project, negotiating resource allocation between needs, and a program plan summary and assignments. The results of this planning process are used to design new studies and test programs, including production and review of statements-of-work and identification of data system needs. This results in implementation and management of studies and test programs through work assignments, QC specifications for each project, and database needs. Each study or test program is monitored for performance and cost-effectiveness.

The long-term and annual planning results are also used to inform data acquisition through contracting for data. Once data are obtained, the data management process consists of inspection of data with regard to conformance to work assignment specifications, data quality inspection, and processing and loading of emissions and non-emissions observations. Tools used for this include data loading programs with automated QC functions. After initial data management and loading, the data is delivered to the MSOD, where test programs are checked for completeness of vehicle descriptive information and tests results. All sources associated with each work assignment are also checked for completeness before being delivered into the MSOD. Once within the MSOD, the data are promoted and distributed via web, compact disc, or FTP. User support is also undertaken, including training, answering questions, and update of support documentation. A formal process and outcome evaluation of each test program is then undertaken.

The databases within MOVES are populated by data from the MSOD.

Determination of Data Sufficiency within MOVES

MOVES is designed to be flexible enough to allow users to provide their own data or make their own judgments as to which sources of data need to be included. For instance, some MOVES users may want to include continuous (1-Hz) emissions data collected from state inspection and maintenance (IM) programs, while others may wish to exclude these data if the data were collected in a location with vastly different meteorology and altitude. Alternatively, a user may wish to provide their own emission rate data (subject to acceptability criteria). MOVES will provide the ability to determine whether such custom applications constitute an "acceptable" application of data. This function will be carried out by the automated MOVES emission data-binning program, which will assign each point in a data set to a VSP bin within each source bin.

Within each VSP bin, the emission data binning program will calculate indices of data quality. These indices are currently under development. Several indices have been proposed, including ratio of standard error to the mean emission rate within each VSP bin and the proportion of within-VSP bin variance contributed by each vehicle. MOVES will automatically flag data that does meet the specified criteria, for example, if 90% of the variance within one VSP bin is contributed by a single vehicle, the data binning program could produce a quality flag indicating caution in interpreting any calculations performed using that VSP bin.

An alternative, as discussed above, is to supplement emission rates from measured data with PERE, the physically-based emissions model. Such a process is required for predicting emissions from future technology vehicles. Furthermore, this possibility allows PERE to supplement existing emissions data in situations where emissions data is difficult to collect and

tends toward high variability (e.g. high-speed, high acceleration events). Subject to quantification of the uncertainty in PERE, this approach offers the possibility of supplementing "real-world" data with "low quality" with modeled data with quantifiable variability and uncertainty. In the context of the emission rate database for the model, PERE can be used to populate source or operating mode bins in the emission rate database for which data of insufficient quantity or quality exist.

MOVES Data Needs Determination

As stated earlier, MOVES design and software implementation is an iterative process, with the highest-priority use cases addressed first. Similarly, as MOVES becomes a finished model, data acquisition will become an iterative process with the model. Specifically, quantitative sensitivity and uncertainty analysis will be undertaken to determine model parameters that make the greatest contribution to overall uncertainty. This process will initially be undertaken primarily by using analytical propagation of error, but future versions of the model may allow for Monte Carlo methods to be implemented where the assumption of normality and sample size sufficiency are not fully met. Model parameters that make the largest contribution to overall uncertainty will be assessed and assigned the greatest weight for data acquisition.

MOVES Quality Assurance Project Plan

MOVES is the first OTAQ model to make formal use of a systematic planning framework. At the time of this document's publication, the first draft QAPP for MOVES is underway. The QAPP will conform to the draft guidelines set forth in OEI's guidance for QAPPs for models. The QAPP will address, in a formal way, project organization and schedule, data quality objectives, staff training requirements, documentation and record-keeping plans, data acquisition needs, data management plans, hardware and software configuration, model assessment activities including sensitivity and uncertainty analysis, software quality assurance, including testing and benchmarking, peer review and management report plans, and data validation.

Conclusion

MOVES is EPA's next generation mobile source emissions model. It has been designed to comply with the suggestions of the NRC panel that reviewed OTAQ's overall modeling framework. Its theoretical foundations, design, data acquisition and management processes are based on the most current understanding of mobile source emissions science, software design, programming, and database management. MOVES design and emissions analysis plans have been subject to independent external peer review, in addition to stakeholder review through the FACA modeling workgroup.

MOVES is the most sophisticated and mature mobile source emissions model that OTAQ has ever developed. The attention to quality planning throughout its development process are producing a model with a high level of utility to decision makers, credibility in the scientific community, and confidence to the model team that we are producing a superior model. Further information about MOVES can be accessed at http://www.epa.gov/otaq/ngm.htm.

NHEERL QA Intranet Page at the MED Link.

[i] National Research Council. Modeling Mobile Source Emissions. Washington: National Academy Press, 2000.

[ii] Environmental Protection Agency. (2002a) Draft design and implementation Plan for EPA's Multi-Scale Motor Vehicle and Equipment Emission System (MOVES). October 2002 peer review draft. [Available at www.epa.gov/otaq/ngm.htm]

[iii] Environmental Protection Agency. (2002b) Draft Emission Analysis Plan for MOVES GHG. November 2002 peer review draft. [Available at www.epa.gov/otaq/ngm.htm]

[iv] See Reference 3.

[v] VSP can be defined by the following equation, adopted from Jose Jimenez at Massachusetts Institute of Technology: $\text{VSP (kW/ton)} = v\left[1.1a + 9.81\left(a\tan\left(\sin\left(grade\right)\right)\right) + 0.132\right] + 0.000302v^3$

[vi] Information and publications on the Comprehensive Modal Emissions Model are available at http://www.cert.ucr.edu/research/project.asp?project=89.

[vii] Environmental Protection Agency. (2000) Mobile Source Observation Data: user guide and reference. Document EPA420-B-00-002.

Systematic Planning for the Residual Risk Program: Adapting the Quality Planning Process to the Needs of the Participants

Malcolm Bertoni and Kara Morgan

The U.S. Environmental Protection Agency requires systematic planning for its programs that create or use environmental data. The basic elements of systematic planning are set forth in the EPA Quality Manual (EPA Manual 5360 A1, May 2000). For many years, the Data Quality Objectives (DQO) process has been the recommended systematic planning methodology, particularly when the application involves decision making (i.e., a clear choice between two or more alternative actions). In the past year, EPA has put forth a draft alternative systematic planning method in Guidance on Systematic Planning for Environmental Data Collection Using Performance and Acceptance Criteria (QA/G-4A), which addresses a broader range of applications using the same scientific principles upon which the DQO process is built. Both the DQO process and the more recent generalization are structured planning methodologies that have well-defined steps, involving numerous stakeholders, which are designed to produce outputs that feed into a Quality Assurance Project Plan (another requirement of the EPA Quality System). While these methods are sound and robust, the authors have experienced many situations where the participants in the planning process get confused about terminology and the purpose of various planning steps, and even question the value of the planning exercise. We believe that most of these stumbling blocks can be avoided if the quality assurance representative adapts the presentation and implementation of systematic planning to the needs of the participants. This paper describes an example of how we adapted the systematic planning process to the needs of the Residual Risk Program within EPA's Office of Air Quality Planning and Standards.

The Residual Risk Program, which is authorized by Section 112(f) of the Clean Air Act, addresses the health and environmental risks due to emissions of hazardous air pollutants (HAPs) that remain after implementation of the Maximum Achievable Control Technology (MACT) standards. Industries that emit HAPs are divided into over a hundred different source categories (such as dry cleaners, coke ovens, halogenated solvents, aerospace, etc.). There is tremendous variety among the source categories in terms of the numbers and locations of emitting facilities, the amounts of annual emissions, the types of releases, the engineering controls that are feasible, and the economic impacts of further controls. Moreover, there is great variety in the quantity and quality of data and other information available to estimate residual risks, evaluate potential controls costs, and estimate economic impacts. Risk assessors, engineers, and economists work together to evaluate whether additional regulations are needed for a given source category, and they often must gather data from a patchwork of different sources. Characterizing the representativeness and uncertainty of their resulting estimates is challenging under the best of circumstances, and they have been embarking on a program streamlining effort to accomplish their mission more efficiently. Under these requirements and constraints, they recognized the need to adopt a systematic planning approach to improve their data collection.

This paper describes the program-wide planning process that was conducted, and shows some of the resulting planning tools that are being used for specific source categories. We supported their planning efforts by first clarifying their current Residual Risk

decision process through an interdisciplinary program-wide planning workshop. Then, we worked with the program team to develop a program-specific planning process that addressed all of the required elements of systematic planning, yet used specific steps and terminology that were more familiar and comfortable for the participants. This program-specific planning process and associated planning tools will be used by source category teams that will do the analysis leading to decisions about whether or not to regulate given source categories. The tools allow the teams to address representativeness and uncertainty using traditional (and more rigorous) quantitative methods as well as innovative qualitative methods. By merging the quality planning steps with their existing planning processes, they were able to obtain agreement on issues that had been unresolved for years.

Practical Estimators of the $PM_{2.5}$ Data Quality Objective Parameters

Basil Coutant, Battelle

Fine particulate matter, particulate matter of aerodynamic diameter 2.5 μm or less ($PM_{2.5}$), is regularly sampled at more than 1,000 sites by state and local EPA agencies across the U.S. for compliance with the National Ambient Air Quality Standards (NAAQS). The Data Quality Objectives (DQOs) for this data collection are generally based on the DQOs developed by the U.S. EPA in 1999 and revised in 2000. The DQOs were established using a simulation model that has seven parameters: sampling frequency, data completeness, precision, bias, seasonality ratio, natural variation, and autocorrelation. The first four of these parameters are controllable through network operations management. The last three are characteristics of the ambient conditions and, hence, are not directly controllable and must be estimated. These same parameters will be required for the forthcoming coarse particulate standard.

In support of these activities, the U.S. EPA has funded the creation of software tools that use the simulation model to develop decision performance curves (also called power curves). However, to use the software, estimates of the ambient characteristics are required as inputs. The goal here is to present practical methods for estimating the input parameters for these tools. The algorithms are derived to establish their technical assumptions and are compared with the output from the simulation model used for the $PM_{2.5}$ DQOs.

1.0 INTRODUCTION

The National Ambient Air Quality Standards (NAAQS) for fine particulate matter, $PM_{2.5}$, has two components (see 40CFR50) based on data collected from three consecutive years. The first component is that the average of the three annual mean concentrations is to be no more than 15.0 $\mu g/m^3$. The second is that the mean of the 98th percentiles for each year is to be no more than 65 $\mu g/m^3$. A nationwide network of over 1,000 sites has been established to monitor $PM_{2.5}$ concentrations providing data for attainment/non-attainment decisions with respect to these two standards. The sampling process is a fairly involved manual process. As a result, for many areas, it is only practical to sample every sixth or every third day, and data completeness is an issue for many sites (see 2000 QA Report). Hence, data quality and the quality of the attainment/non-attainment decisions made based on the data are being carefully monitored by the U.S. EPA (EPA).

EPA uses the Data Quality Objective (DQO) process (see EPA's guidance document QA/G-4). This process quantifies the data quality needs so that the subsequent decisions meet decision maker needs. As a result of the application of that process, a software tool, *DQO Companion*, was created for EPA. The tool allows the local agencies to tailor the DQO process to the local ambient behavior of $PM_{2.5}$. (See U.S. EPA's DQO *Companion*, Version 1.0 User's Guide.) As the tool demonstrates, the local behavior can have a substantial effect on the quality of the attainment/non-attainment decision. However, in order to use the tool, certain parameters need to be estimated from local data. This paper describes fairly easy and robust methods of obtaining estimates of the seasonality ratio, population Coefficient of Variation (CV), and autocorrelation.

2.0 THE SIMULATION MODEL

The simulation models both the truth and the sample values. Truth is based on a sinusoidal curve with a period of one year with a vertical shift that is constrained to keep the curve strictly above 0. This curve represents the long-term seasonal mean. The idea is to mimic the oscillatory behavior associated with seasonal changes. Simulations with different shaped curves have shown that the power curves are not sensitive to the explicit shape, but are sensitive to the extremes. Hence, the curve is parameterized by a long-term annual mean (the vertical shift from 0) and the ratio of the minimum of the curve to the maximum of the curve. Next, day-to-day deviations from the curve are applied multiplicatively. These deviations are assumed to have a lognormal distribution with a mean given by the sine curve and a constant coefficient of variation (CV). This population CV is assumed to be a property of the local conditions at a site. The true deviations are allowed to be correlated on the log scale. As with the CV, the magnitude of any correlation is assumed to be site-specific.

The simulations then apply a multiplicative bias, random measurement error (independent, normally distributed with a constant measurement CV), a fixed sampling schedule of one sample every m days, and random missing data with a specific quarterly completeness. Finally, the power curves plot the probability of observing a 3-year aggregate value greater than the standard versus the true 3-year aggregate value. Note that the true 3-year aggregate value is the realization of a random process in this case, not the more usual long-term expected value. The probabilities are based on simulating at least 5,000 3-year realizations of both truth and measured values.

3.0 ESTIMATED SIMULATION PARAMETERS

The subsections of Section 3 describe the parameters that control the truth in the simulations. The three parameters are the seasonality ratio, the population CV, and autocorrelation. All the recommended procedures require at least one year of data. It is assumed that national level or region level DQOs would be used for at least the first year. The current network has been operational for three years, so there are ample data to obtain the initial estimates.

Each of the recommended estimators has been tested against the simulation model with a long-term mean of approximately15.0 $\mu g/m^3$. Of the three estimators, the autocorrelation estimator is the most variable, and it is slightly biased high (probably due to the constraint that it be non-negative).

3.1 Seasonality Ratio

The ratio parameter is a measure of the degree of seasonality in the data. In the model, it is the ratio of the high point to the low point on the sine curve. The model assumes that the amplitude of the sine curve is proportional to the mean. Hence, it is assumed that ambient concentrations will be proportionally reduced throughout the year. This is felt to be approximately true for areas of concern but has not been investigated.

The estimator recommended is based on finding monthly mean concentrations. The estimate is the ratio of the highest monthly mean divided by the lowest monthly mean. Where more than a year of data are available, the data from each month are combined across years for a total of 12 monthly means. Bi-monthly means have been tested and, as expected, reduce the extremes. However, usually the difference is negligible.

3.2 Population CV

This parameter measures the amount of random, day-to-day variation of the true concentration about the sine curve. The procedure for estimating this parameter is a bit more involved than the estimate for the seasonality ratio estimate. Also, since it is a variance parameter, it is harder to estimate in that it requires more data to achieve a comparable level of precision. The following does a reasonable job.

Start with every 6th day measurements (deleting, if needed) and take the natural log of each. Every sixth day measurements are used to avoid the effects of autocorrelation. The logarithm is used to match the model, namely that a log-normal distribution is used for generating the deviations from the sine curve. Create a new sequence of numbers equal to the differences of successive pairs in the sequence of the logs. Ignoring measurement error for the moment, these differences will be of the form:

$$\ln(\text{Seasonal trend at day } i+6) + \ln(\text{random deviation for day } i+6) - [\ln(\text{Seasonal trend at day } i) + \ln(\text{random deviation for day } i)]$$

$$= \ln(\text{Seasonal trend at day } i+6) - \ln(\text{Seasonal trend at day } i) + \ln(\text{random deviation for day } i+6) - \ln(\text{random deviation for day } i)$$

$$\approx \ln(\text{random deviation for day } i+6) - \ln(\text{random deviation for day } i).$$

The measurement error should be less than 10 percent and, hence, much smaller than the random component of the day-to-day deviations, which is typically 50 percent or more.

The next step is to remove every other term from the sequence generated above so that the terms are mutually independent. Now, assume that the random deviations from the seasonal trend are generated from a log-normal distribution with a mean of 1 and a geometric variance of σ^2. Then, the variance of the terms of the sequence generated is $2\sigma^2$. Hence, the geometric variance of the random deviations about the seasonal trend can be estimated by one-half the variance of the sequence generated above. Or, equivalently, the CV of the deviations can be estimated by $\sqrt{\left(\exp\left(S^2/2\right)-1\right)}$, where S is the standard deviation of the sequence generated above.

3.3 Autocorrelation

The final parameter describing the natural variation of the true concentrations is the autocorrelation of the deviations from the seasonal trend. This is a measurement of the similarity between successive days. Consider two sets of measurements. First, suppose one set of measurements contains the measured $PM_{2.5}$ concentration on each July 15th for the past five years. The population CV is intended to capture how different these measurements are from each other. On the other hand, suppose the second set of measurements have the $PM_{2.5}$

concentrations for each day from July 15, 2002, to July 20, 2002. These measurements may not be as spread out as the other values simply because they were taken closer together in time. Autocorrelation measures this effect. A good way to think of autocorrelation is that it measures how quickly the local concentrations can change. The value of the autocorrelation under consideration is between 0 and 1. A value of 0 indicates that the local concentrations change very fast. A value of 1 indicates that the local concentrations are constant.

Estimating autocorrelation is even harder than estimating the population CV. Without daily observations, it is recommended that 0 be used in the DQO tool because 0 is the most conservative value. In fact, for most sites considered, any autocorrelation present is small, so 0 is still a reasonable value to use in the DQO tool.

Assume that daily measurements are available. Then let $S6$ be the standard deviation computed in the section on population CV, based on differences of the logs from every 6th day measurements. Let $S1$ be the corresponding standard deviation based on differences of the logarithms of the concentrations from successive daily measurements. If $S6 > S1$, then there is evidence of some autocorrelation.

The differences from the every sixth day measurements can be taken to be effectively independent since the correlation between these measurements should be the same as that of the daily measurements to the sixth power. Hence, the variance of the quantity:

ln(random deviation for day i+6) - ln(random deviation for day i)

is $2\sigma^2$, where σ^2 is the geometric variance of the random deviations about the seasonal trend as before. However, under the assumption of correlation, the variance of the quantity:

ln(random deviation for day i+1) - ln(random deviation for day i)

is $2\sigma^2 - 2\rho\sigma^2$, where σ^2 is the geometric variance of the random deviations about the seasonal trend. Hence, $\left(S6^2 - S1^2\right)/S6^2$ is an estimate of the autocorrelation.

Testing with the simulation model has shown that this estimate can be variable, as should be expected since it is an estimate of a variance parameter for a non-linear model. Since it is better to underestimate this parameter (to make the results more conservative) the DQO tool manual suggests multiplying the estimate by 0.85.

4.0 DISCUSSION

The primary criterion for choosing the three estimators was that they be easy to apply, in that they could easily be calculated in a spread sheet. The three estimators satisfy that criteria. They also do a quite reasonable job when based on at least one year of data. However, while the estimators were developed for site-specific estimates, they probably should not be applied to a site in isolation for DQO development. DQOs are generally applied across a group of sites, usually no smaller than a reporting organization. Hence, as was done for the national level DQOs, the range of site-specific parameter estimates should be considered. In particular, upper

bounds are needed for the population CV and the seasonality ratio across all sites. A lower bound on the autocorrelation produces the "worst case," the most difficult case for measuring the annual mean and the 98th percentile. These extremes should be used in the DQO tool to establish the measurement precision and bias levels that will meet decision maker needs.

4.0 REFERENCES

Cassella, George, and Berger, Roger, (1990). Statistical Inference, Wadsworth Publishing, Belmont, California.

Code of Federal Regulations (July 1997). "40CFR50," downloaded from http://www.access.gpo.gov/nara/cfr/index.html accessed November 6, 2002.

U.S. EPA (2000). "Guidance for the Data Quality Objectives Process (EPA QA/G-4)." Report No. EPA/600/R-96/055, August.

U.S. EPA (2002). "DQO *Companion*, Version 1.0 User's Guide," written by Battelle for U.S. EPA under Contract No. 68-D-98-030, Work Assignment 5-07.

Evaluation and Use of Secondary Data

James T. Markwiese, Neptune and Company

All scientific disciplines rely upon existing (secondary) data to design new studies, test hypotheses and to make decisions. An understanding of secondary data quality is useful for these efforts. For example, existing toxicological dose-response data are often used to determine levels of contaminants in environmental media that will require remediation. Because any record can represent secondary data, a framework for addressing the existing data quality must be general yet comprehensive. Secondary data will be evaluated according to representativeness and comparability and other applicable data quality indicators. Guidelines are presented for collecting and evaluating the acceptability of secondary data. Specifically, a two-stage screening process is described and attendant data quality considerations are presented for each stage in the context of addressing project objectives.

Introduction. This presentation describes a transparent and objective means of identifying relevant information as well as selecting key data from that information to address a problem statement. An ecological risk assessment example is used to illustrate this process. Specifically, we are interested in defining a safe concentration of a contaminant in the environment; i.e., a concentration below which adverse effects are not expected.

State the Problem. The first step in any systematic information gathering effort is to clearly define the issue to be addressed (EPA 2000a). This task corresponds to the first step in the Data Quality Objectives (DQO) process; i.e., state the problem (Figure 1). A succinct and comprehensive statement of the issue facilitates the establishment of secondary data acceptance criteria suitable for the project's objective. These criteria are described in more detail in the screening step for identifying key data for addressing the project objective.

For purposes of illustration, consider the problem stated as defining a no adverse effect concentration (NOEC) for cadmium in the environment using existing information. A search based on this loosely defined objective (using search terms of cadmium, NOEC and environment) could generate a tremendous number of references. The assessment of applicable information among this vast literature would not be straightforward. Given the wealth of information potentially available to the environmental scientist, it is necessary to focus on only those secondary data that directly pertain to the problem statement.

A more specific objective, e.g., defining a concentration of cadmium in soil that is associated with no observed adverse effects in plants over chronic exposure, would narrow the number of references. A clear statement of the issue allows for an evaluation of whether the secondary data are representative of project needs. Representativeness is defined as, "the measure of the degree to which data accurately and precisely represent a characteristic of a population, parameter variations at a sampling point, a process condition, or an environmental condition" (ANSI/ASQC 1994). This definition captures multiple scales of information. In the example of cadmium toxicity to plants, data users are in a much better position to focus on representative information

with a better-defined problem statement. For example, an article entitled, "cadmium toxicity to wheat" is worthy of further scrutiny whereas an article describing the role of cadmium exposure in fish gill lesions would be eliminated from consideration.

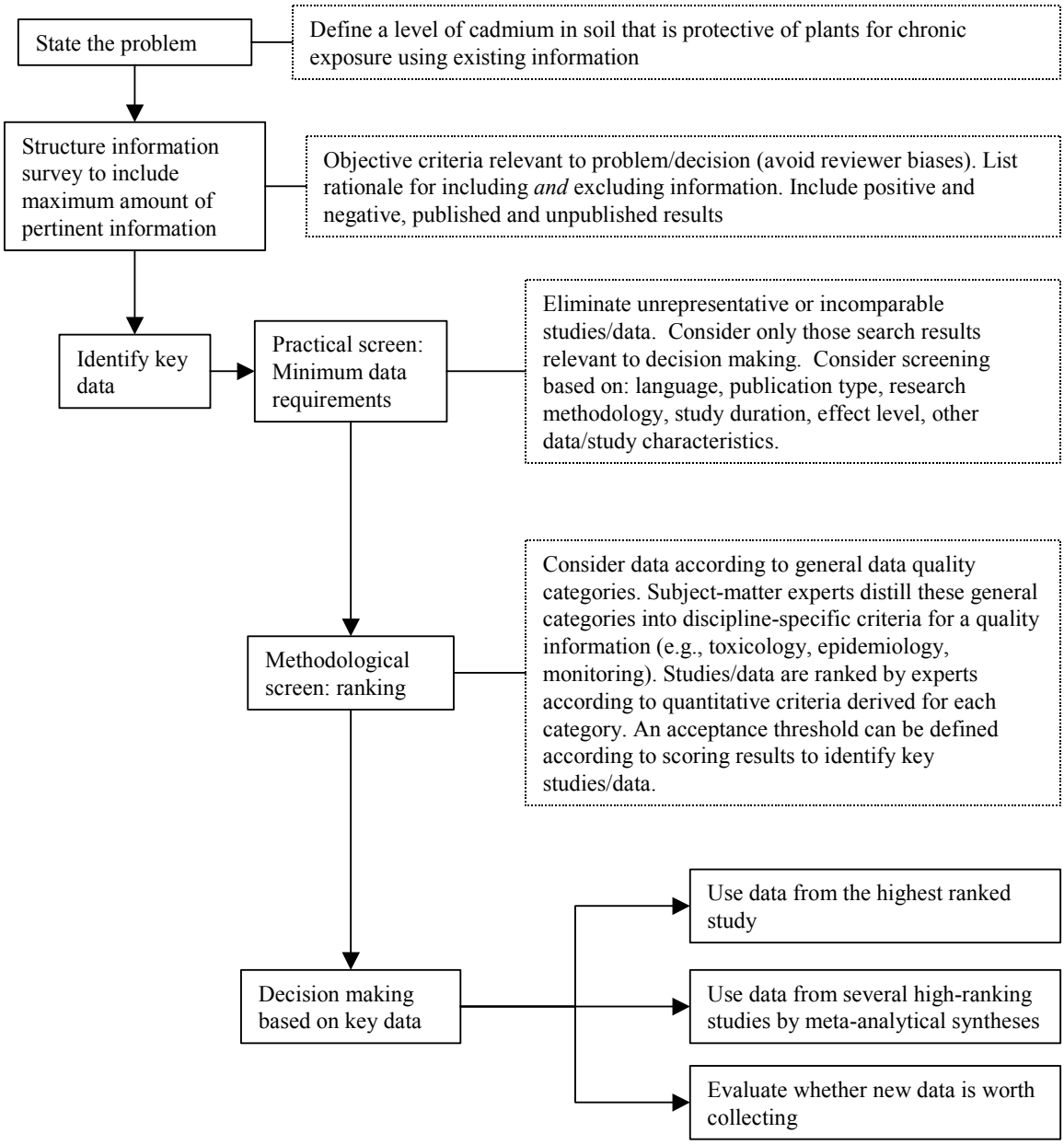

Figure 1. A process for collecting and analyzing secondary data to address the problem statement

Information Survey. Upon establishing the problem statement (Figure 1), a survey can be structured to capture the maximum amount of pertinent information (Fink 1998). It is important to explicitly list the rationale for information search parameters (WHO 2000). This helps safeguard against reviewer biases influencing the data set to be critiqued because reviewers may

consciously or unconsciously include information that favor their own biases while ignoring that which does not (Slavin 1995).

The sources of information consulted will depend on the study objective and could include published literature (both in peer-reviewed journals and so-called grey literature); relevant databases; and, predictions from models. A comprehensive assessment of available data helps avoid biases. For example, sole reliance on published literature may result in an overstatement of effects of a treatment because it is more difficult to publish studies showing no effects (Slavin 1995). Depending on the importance and scope of your project, one would set up a more or less elaborate review. This "graded approach" provides flexibility in that the ultimate rigor in evaluating data is dictated by the intended use of the data (EPA 2000a)

Identification of Key Data. The information survey designed to address the problem statement will likely result in a considerable amount of information. To most efficiently sieve through this material, it is recommended that a two-stage screening procedure be employed (Fink 1998, EPA 1999, EPA 2000b, Fennel 2002). A priori criteria are established to evaluate data quality in these screening stages. Initially, the data are evaluated according to whether minimum requirements, such as representativeness and comparability, are sufficient to directly address the project objective. For example, EPA's High Production Volume Challenge Program requires information on the route and type of stressor exposure before data sets can be further evaluated for ecotoxicity and human health effect endpoints (EPA 1999). Considering the example problem statement, only cadmium-plant toxicity data using soil as an exposure pathway would be suitable for our purposes. Similarly, EPA's Ecological Soil Screening Levels (Eco-SSLs) process requires meeting eleven literature-acceptance criteria to advance to the next stage of data abstraction (EPA 2000c). Under the Eco-SSL criterion of toxicant, only data with adverse effects caused by a single stressor can advance (with multiple stressors, it is difficult to attribute causality of observed adverse effects, such as mortality, to any one stressor).

Studies or data meeting minimum requirements proceed to a second stage (Fink 1998, EPA 1999, EPA 2000d, Fennel 2002). This is referred to as the methodological screen (Figure 2). The methodological screen more specifically addresses issues of data adequacy. For data quality categories, subject matter experts quantitatively score data. Many authors have posited such categories, including reliability, validity and accuracy (Fink 1998); reliability, relevance and adequacy (Klimisch et al. 1997); and, soundness, applicability and utility, clarity and completeness uncertainty and variability, evaluation and review (EPA 2002).

As illustrated by these categories, data quality characteristics must be comprehensive enough to apply to the many potential sources of information available to an investigator (consider the inherent differences among monitoring, survey and experimental data). By nature of this inclusiveness, however, the categories are necessarily broad. Subject matter experts develop more specific criteria from general categories. For purposes of illustration, data quality assessment factors proposed by the Office of Environmental Information (EPA 2002) are developed into checklist examples to address the objective of defining a concentration of cadmium in soil that is protective of plants over chronic exposures (Figure 2).

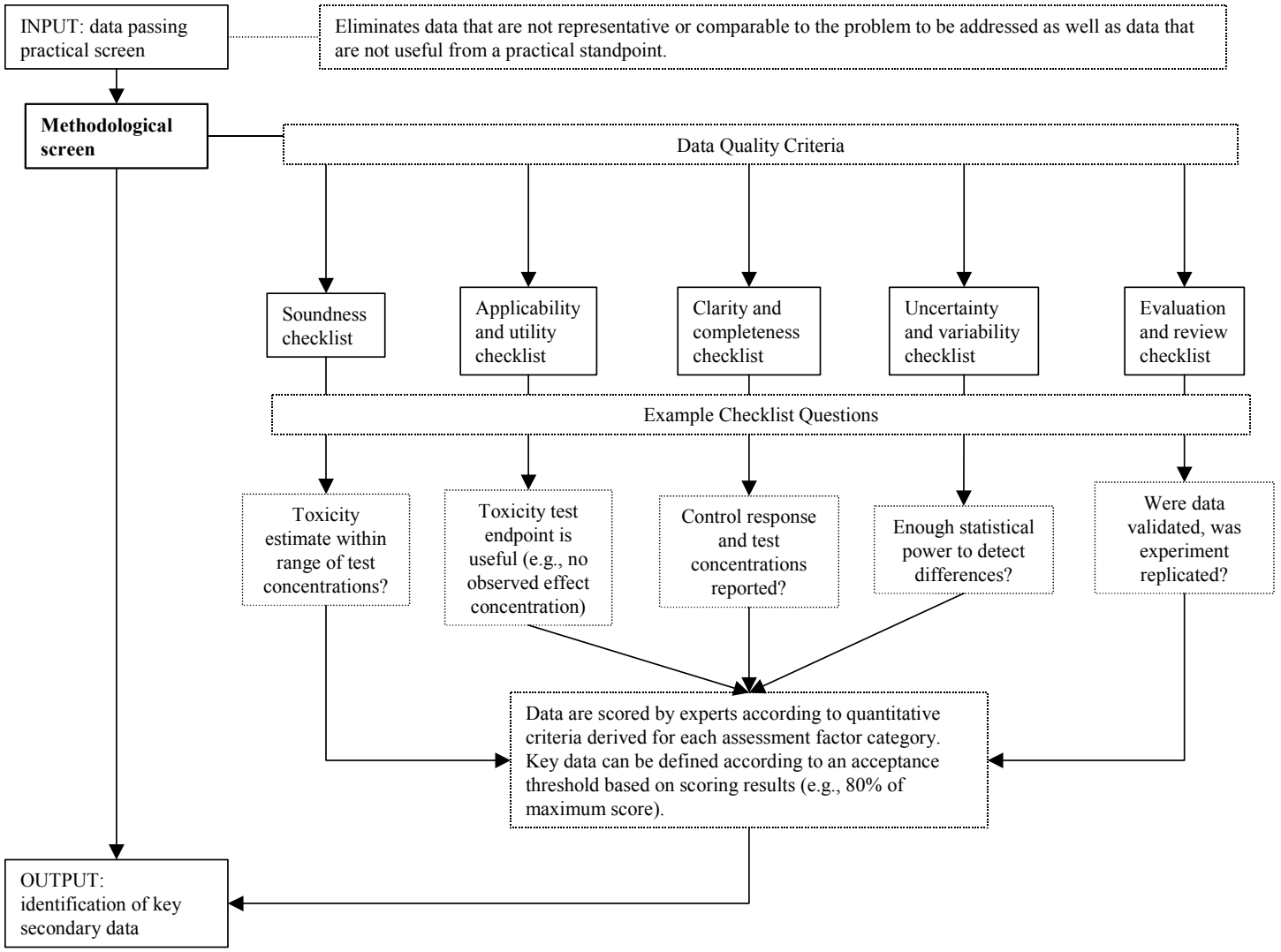

Figure 2. Input from the practical screen is quantitatively ranked in the methodological screen to identify key data from the information collected. EPA's Office of Environmental Information data assessment factors (EPA 2002) are used as example data quality categories.

Checklists are used to quantitatively rank data (EPA 2000d). Considering the Eco-SSL protocol, the second stage screen employs nine criteria to assess the applicability of data for deriving toxicity thresholds and provides a set of rules for extracting and reporting the most appropriate data. For each criterion, scoring is based on a three-point scale: 0, 1 or 2, with 2 being the highest score indicating complete agreement with the criterion.

In the criterion to determine adequacy of test duration, for example, the data receive a score of "2" if chronic exposures or life-cycle phase studies were used, a score of "1" if acute tests were used and a score of "0" if very short-term exposures were used. The scores for each criterion are recorded and summed to generate a total score for each data set. Key data can be quantitatively defined according to an acceptance threshold based on scoring results (e.g., 80% of maximum score). This establishment of an acceptability threshold for decisions based on secondary data is defensible in that justification is provided for selecting the data best suited to address the project objective.

Decision Making Based on Key Data. Consider again the project objective of defining a concentration of cadmium in soil that is protective of plants over chronic exposure. To address such an objective, governmental organizations like Oak Ridge National Laboratory (Efroymson et al. 1997) and Los Alamos National Laboratory (LANL 2003) have typically based toxicity thresholds on the data from the best available or critical study. In contrast, the recent Superfund guidance for ecological soil screening levels (EPA 2000b) uses several of the highest-scoring data sets to define safe levels of chemicals in the environment. The latter process maximizes the use of the available data for deriving soil ecological screening levels.

Regardless of the decision-making process to define "safe" levels of chemicals in the environment, the approach described herein provides for a transparent and objective means of identifying relevant information as well as identifying key data from that information. Basically, this approach illustrates a means of employing defensible data in decision making (Figure 1). Once the review is completed and the data are ranked, the scientific investigator is in a position to determine whether the data are adequate to address the problem statement or whether new data need to be collected.

References

ANSI/ASQC E4 (1994) Specifications and Guidelines for Environmental Data Collection and Environmental Technology Program. American Society for Quality. Milwaukee, WI.

Efroymson RA, Will ME, Suter GW II, Wooten AC (1997) Toxicological Benchmarks for Screening Contaminants of Potential Concern for Effects on Terrestrial Plants: 1997 Revision. Oak Ridge National Laboratory, Oak Ridge, TN. 128 pp. ES/ER/TM-85/R3. http://www.hsrd.ornl.gov/ecorisk/tm85r3.pdf

EPA (1999) Determining the Adequacy of Existing Data. US Environmental Protection Agency, High Production Volume (HPV) Challenge Program http://www.epa.gov/chemrtk/datadeqfn.pdf

EPA (2000a) Using Data from Other Sources – A Checklist for Quality Concerns, US Environmental Protection Agency, EPA Quality Staff, http://www.epa.gov/quality1/qs-docs/cklist-secondary.pdf

EPA (2000b) Ecological soil screening level guidance – draft. US Environmental Protection Agency http://www.epa.gov/superfund/programs/risk/ecorisk/guidance.pdf

EPA (2000c) Ecological soil screening level guidance – draft, Exhibit 3-1, Plant and soil invertebrates standard operating procedure #1: literature search and acquisition. Attachment J, Literature acceptance criteria checklist, US Environmental Protection Agency http://www.epa.gov/superfund/programs/risk/ecorisk/exhibits/exhibit5.pdf

EPA (2000d) Ecological soil screening level guidance – draft, Appendix 3-1, Plant and soil invertebrates standard operating procedure #3: literature evaluation and data extraction. US Environmental Protection Agency http://www.epa.gov/superfund/programs/risk/ecorisk/appendix.pdf

EPA (2002) Assessment factors for evaluating the quality of information from external sources, Draft. US Environmental Protection Agency, Office of Environmental Information, http://www.epa.gov/oei/qualityguidelines/af_assessdraft.pdf

Fennell DB (2002) A white paper on compiling and evaluating research information for secondary use in ORD's assessment documents [draft]. Research Triangle Park, NC: U.S. Environmental Protection Agency, National Center for Environmental Assessment-RTP Division; January 7.

Fink A (1998) Conducting research literature reviews: from paper to the internet. Thousand Oaks, CA: Sage Publications, Inc.

Klimisch HJ, Andrae M, Tillmann U (1997) A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regulatory Toxicology and Pharmacology 25:1-5.

LANL (2003) ECORISK Database (Release 1.5), ER package #186, Environmental Restoration Project, Los Alamos National Laboratory, Los Alamos, New Mexico.

Slavin RE (1995) Best evidence synthesis: an intelligent alternative to meta analysis. Journal of Clinical Epidemiology 48:9-18.

WHO (2000) Evaluation and use of epidemiological evidence for environmental risk assessment: guidance document. World Health Organization, Copenhagen, Denmark: Regional Office for Europe.

An Information Quality Life Cycle Model: Extending the Reach of Quality Management to Secondary Users of Data

Malcolm Bertoni, Ian Beaty, Kara Morgan, Cynthia Salmons, Robert S. Wright

The U.S. Environmental Protection Agency's quality system has employed a data quality life cycle perspective for many years. Planning, implementation, and assessment are the three main stages in the data quality life cycle, which form the foundation upon which numerous quality procedures and tools are built. Until recent years, the EPA quality system had applied this life cycle framework primarily to the generation of new environmental measurement data. However, for many reasons, the focus of the EPA quality system has been expanding beyond measurement data to include a variety of other sources, such as modeling results, scientific literature, historical databases, and surveys, to name a few. Moreover, the explosion of information technology and emergence of the world wide web has made data and information widely available for secondary uses that may not have been contemplated by those who originally created the data.

Over the past year, EPA and other federal agencies and departments have addressed concerns about information quality through the development of guidelines mandated by the U.S. Office of Management and Budget. Some of the concepts and terminology introduced by OMB are different from what EPA's quality community has developed over the past 30 years. A question that environmental quality assurance practitioners may rightly ask is, "How do these information quality guidelines relate to our traditional data quality assurance program?"

Information quality encompasses the Agency's previous data quality paradigm and extends it to address a more comprehensive view of the "intended use" of data. Just as information can be viewed as "data in context," information quality can be viewed as encompassing yet going beyond traditional data quality indicators (accuracy, completeness, etc.) to include presentation format, timeliness, authenticity, and other quality characteristics that affect how well the information supports a particular need at a required place and time.

This paper presents an information quality life cycle model that illustrates how the traditional planning, implementation, and assessment steps of the data quality life cycle can be extended to address information production, storage, and distribution. Because the model shows how data and information are stored and reused, it helps identify how secondary users of existing data fit into the quality system. The model represents a synthesis of several information quality perspectives from the leading researchers and authors in that field, as well as several EPA sources. The model attempts to bridge the terminology gap between the information technology community and the environmental science and policy community. The main elements of the life cycle model are described, and several examples are developed to explain how the model can be used to improve quality planning and assessment.

A Win-Win-Win Partnership for Training Environmental Statisticians - A Panel Discussion

Dr. Kimberly Weems and Bill Hunt from NC State University; Dr. Nagambal Shah and Dr. Monica Stephens, Spelman College; Michael Crotty, NCSU graduate student; Barry Nussbaum and Cary Roberts, U.S. EPA; David Mintz, U.S. EPA; Steve Few, North Carolina Department of Environment and Natural Resources; Van Shrieves, U.S. EPA

How could a win-win-win strategy be used to train young people in environmental statistics, simultaneously analyze environmental data for Federal, State and local agencies and increase the number of undergraduate students going on for advanced degrees in statistics? Two courses have been developed to train undergraduate students in environmental statistics and provide them with a consulting experience. Bill Hunt, Visiting Senior Lecturer in the Department of Statistics at North Carolina State University, developed the courses. He has over 35 years experience in working on the analysis of environmental data and using important results to make national air pollution policy decisions. He has been active in the ASA and served as chair of the Environmental Section. During his 35 year career with the USEPA, he has tried on numerous occasions to increase the analysis of environmental data. Many efforts were made to encourage the hiring of statisticians and data analysts to analyze environmental data at the Federal, State and local levels, but all of these efforts were met with very limited success. The government continues to collect data, which largely go unanalyzed.

The environmental courses developed at NC State University are intended to make a dent in this problem and reinvigorate the relationship between universities and colleges and environmental agencies. Along with his colleague, Dr. Kimberly Weems, they will be working with Dr. Nagambal Shah and Dr. Monica Stevens of Spelman College and Mr. Van Shrieves of EPA Region 4 to reproduce this program at Spelman College. Spelman College will work with the U. S. Environmental Protection Agency Region 4 Office in Atlanta, GA and the Georgia Department of Natural Resources. This collaborative effort will demonstrate that their approach can be used at universities and colleges with an undergraduate statistics program and at those without, as long as there are some courses in statistics.

The objectives of the environmental statistics courses are: (1) to provide a consulting opportunity for the students with Federal, State or local agencies; (2) focus on the application of the student's technical skills to a real problem; (3) have the students gain consulting experience; and (4) develop their oral and written communication skills. The students learn how to prepare a final report, brief clients at the client's office, and present poster papers at technical conferences and write papers for publication. Students have done work for eight clients: (1) the Southern Oxidant Study at North Carolina State University (NCSU); (2) the U. S. Environmental Protection Agency's (USEPA) National Exposure Research Laboratory; (3) the USEPA's Office of Air Quality Planning and Standards; (4) the North Carolina Department of Environment and Natural Resources (NCDENR); (5) the Forsyth County Environmental Affairs Department; (6) the U. S. Department of State; (7) the USEPA's Office of Environmental Information in Washington, DC ; (8) Environment Canada; (9) Texas Commission on Environmental Quality; and the (10) University of Texas.

Bill Hunt and Kimberly Weems will provide an overview of the course. They will comment on the impact the course has had on the university - more students pursuing graduate study, the faculty develops new contacts with environmental agencies, and the placement of their students in rewarding careers. The USEPA panelists can discuss the usefulness of having their data analyzed, the impact of student analyses on environmental policy decisions, and their experience in hiring the students for permanent or part time work. Mr. Van Shrieves will talk about implementing the course in Region 4. Dr. Nagambal Shah and/or Dr. Monica Stephens will also comment on the implementation of the environmental statistics program in Region 4 from the perspective of Spelman College. Mr. Steve Few can present the State viewpoint to complement the USEPA viewpoint. Michael Crotty can discuss the course from a student's perspective - gaining experience in doing research, consulting, writing reports, giving briefings, presenting papers, etc and how the course helped him in his career choices.

The panel will discuss the question: Can pair universities/colleges (Statistics Departments and Math Departments with statistics courses) be paired with environmental agencies across the United States and make this program work. The audience will be invited to ask questions.